



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 119347

TO: Minh-Tam Davis
Location: rem/3a24/3c18
Art Unit: 1642
Tuesday, April 13, 2004

Case Serial Number: 09/763335

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1A69
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

Search Notes

Tam,
Paul Schulwitz assisted me with the Score/Length search. There were no hits that met your criteria of length between 60-200 nucleotides, and score over length value of 70% or greater. Consequently, there is no print-out for that segment of the search.
Please let me know if you have questions.
Barb

DR P-PSDB; AAY70400.

XX New human cell signaling proteins and polynucleotides useful for
PT diagnosis, prevention and treatment of neoplastic, neurological,
PT immunological, vesicle trafficking and smooth muscle disorders.

XX Claim 9; Page 61; 62pp; English.

XX The present cDNA sequence encodes human cell-signaling protein-2 (CSIG-
CC 2) involved in regulation of cell proliferation, differentiation and
CC gene transcription. The CSIG-2 cDNA was isolated from brain tissue cDNA
CC library (BRINOT12). The BRINOT12 library was constructed using RNA
CC isolated from brain tissue removed from the right frontal lobe of a 5-
CC year old Caucasian male during a hemispherectomy. CSIG cDNA can be used
CC in the diagnosis and treatment of diseases associated with expression of
CC CSIG. These diseases include neoplastic, neurological, immunological,
CC vesicle trafficking and smooth muscle disorders, including HIV,
CC rheumatoid arthritis, asthma, atherosclerosis, diabetes mellitus,
CC emphysema, irritable bowel syndrome, multiple sclerosis, osteoporosis,
CC psoriasis and infections including viral, bacterial and fungal. This is
CC also useful for generating hybridisation probes useful in mapping the
CC naturally occurring genomic sequences

SQ Sequence 812 BP; 243 A; 162 C; 195 G; 212 T; 0 U; 0 Other;

Query Match 100.0%; Score 812; DB 3; Length 812;

Best Local Similarity 100.0%; Pred. No. 2.6e-210;
Matches 812; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CACTGAGTGGGAGTGTCCATCGGCAACTTAACTATTCATCATGAGAACTGCACA 60
DB 1 CACTGAGTGGGAGTGTCCATCGGCAACTTAACTATTCATCATGAGAACTGCACA 60
QY 61 TTATCTCCCATCATTTGAAAGGTGTGTCAGGCAAGGTAACGCAAGATATTTAA 120
DB 61 TTATCTCCCATCATTTGAAAGGTGTGTCAGGCAAGGTAACGCAAGATATTTAA 120
QY 121 AGGTGAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 180
DB 121 AGGTGAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 180
QY 121 AGGTGAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 180
DB 121 AGGTGAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 180
QY 181 TGAATGAACCCGATGCTTTTTTTTACTGTGGAATGAGATCGAAGAGTAACATTT 240
DB 181 TGAATGAACCCGATGCTTTTTTTTACTGTGGAATGAGATCGAAGAGTAACATTT 240
QY 241 TTTTCTTAACTCTGATTAAGAGATGTTGGAGAGCTTTTGAAAAAAATTTAAATG 300
DB 241 TTTTCTTAACTCTGATTAAGAGATGTTGGAGAGCTTTTGAAAAAAATTTAAATG 300
QY 301 TGGCAGATGATTTTAAAAAGTGTAGATCTTCCATGACACTATAGATGACTCT 360
DB 301 TGGCAGATGATTTTAAAAAGTGTAGATCTTCCATGACACTATAGATGACTCT 360
QY 361 GCTCTGGCTGATTTTTCAGAGATGCAATGCTCTGCGAGTCTCTGGGTCTGTAT 420
DB 361 GCTCTGGCTGATTTTTCAGAGATGCAATGCTCTGCGAGTCTCTGGGTCTGTAT 420
QY 421 TTGTGGATTAAGTCTTGTGCAATGCTTACTCTGCAATGATCCCTTCACACATTTTCCAG 480
DB 421 TTGTGGATTAAGTCTTGTGCAATGCTTACTCTGCAATGATCCCTTCACACATTTTCCAG 480
QY 481 CAGCATCCTGTCAGACAGCCAGAGAGAGAGTGTGAAGTGAAGACACACCGATGT 540
DB 481 CAGCATCCTGTCAGACAGCCAGAGAGAGAGTGTGAAGTGAAGACACACCGATGT 540
QY 541 TGTAAACAAGATCGATTTGAGAGCGGTCAACAACATGAAGTGTCTGTCTACTGGA 600
DB 541 TGTAAACAAGATCGATTTGAGAGCGGTCAACAACATGAAGTGTCTGTCTACTGGA 600
QY 601 AAAGTGGCTGGAACAACAAGAAACCGGCTTCTTGTGCTGATGCTCCATAGTATGGG 660
DB 601 AAAGTGGCTGGAACAACAAGAAACCGGCTTCTTGTGCTGATGCTCCATAGTATGGG 660

```

```

QY 661 AATGTGTGTGATGATGAGAGCTTGCTTAGAGAGAGAAATGTAAAGACTCCCTGAC 720
DB 661 AATGTGTGTGATGATGAGAGCTTGCTTAGAGAGAGAAATGTAAAGACTCCCTGAC 720
QY 721 AATTCGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 AATTCGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 ACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812
DB 781 ACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812

```

RESULT 2
AB077548
ID AB077548 standard; cDNA; 2058 BP.

AC AB077548;

XX 01-OCT-2002 (first entry)

DE Human phosphorylcholine/ethanolamine transferase 9.35-encoding cDNA.

XX Human; phosphorylcholine/ethanolamine transferase 9.35;
KW recombinant production; gene therapy; lung hypotasia;
KW bile acid metabolism disorder; angiodysplasia; tumor; cancer;
KW immune disorder; inflammatory condition; cystostatic; antiinflammatory;
KW immunomodulator; gene; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FH CDS 399..656

FT /product= "human phosphorylcholine/ethanolamine

FT transferase 9.35"

XX CN1341736-A.

XX 27-MAR-2002.

XX 05-SEP-2000; 2000CN-00125017.

XX 05-SEP-2000; 2000CN-00125017.

XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-501219/54.

XX P-PSDB; AAM49386.

XX New polypeptide-human phosphorylcholine/ethanolamine transferase 9.35 for

XX treating lung hypotasia, bile acid metabolic disturbance disease,

XX angiodysplasia, related tumor, inflammation and immunological disease.

XX Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.

XX The invention relates to human phosphorylcholine/ethanolamine transferase
CC 9.35 (AAM49386) and nucleic acids encoding it (AB077548). The protein has
CC a molecular weight of 9.35 kD. The invention also relates to a method for
CC the recombinant production of the protein, an antagonist of the protein,
CC and the use of the protein, gene and antagonist in therapeutic
CC applications. Phosphorylcholine/ethanolamine transferase 9.35 can be used
CC in the treatment of a variety of diseases such as lung hypotasia,
CC disorders of bile acid metabolism, angiodysplasia, related tumors,
CC immune disorders and inflammatory conditions. The present sequence
CC represents cDNA encoding human phosphorylcholine/ethanolamine transferase
CC 9.35

XX Sequence 2058 BP; 609 A; 402 C; 437 G; 610 T; 0 U; 0 Other;

Query Match 84.5%; Score 686; DB 6; Length 2058;

Best Local Similarity 89.4%; Pred. No. 1e-192;
Matches 812; Conservative 0; Mismatches 0; Indels 96; Gaps 3;

QY	1	CACTGAGAGTGGGAGATGGTCCATCGGCAACCTTAAATCGATTCTCATCAGAAACTGCA	60
Db	59	CACGTGAGTGGGAGATGGTCCATCGGCAACCTTAAATCGATTCTCATCAGAAACTGCA	118
QY	61	TTATCTCCCATCACTTCMAAGSTCTGTCAGGCAAGGTGACGCCAGAGATGATTAA	120
Db	119	TTATCTCCCATCACTTCMAAGSTCTGTCAGGCAAGGTGACGCCAGAGATGATTAA	178
QY	121	AGGTGAAATGACAAAGTTTCCACCCCTCAACCTTGCTGCTCTTCTGCAATAAGTC	180
Db	179	AGGTGAAATGACAAAGTTTCCACCCCTCAACCTTGCTGCTCTTCTGCAATAAGTC	238
QY	181	TGATGAACCCGATGTCTTTTTTTTAACTGTGGAATATGATCGGAGAGATGATACA	-T 238
Db	239	TGATGAACCCGATGTCTTTTTTTTAACTGTGGAATATGATCGGAGAGATGATACA	TTT 298
QY	239	TTTTTTTTTAACTCGTATTAAGAAGATGTGGAGCTCTTTGAAAAAAATTTTAAAT	298
Db	299	TTTTTTTTTAACTCGTATTAAGAAGATGTGGAGCTCTTTGAAAAAAATTTTAAAT	358
QY	299	TGTGGCAACATATGATTTTAAAAAGTGTAAATCTTTTCAATGAACCTAATAGATACT	418
Db	359	TGTGGCAACATATGATTTTAAAAAGTGTAAATCTTTTCAATGAACCTAATAGATACT	418
QY	359	CTGCTCTGGCTGGATTTTTCAGAGATGGCAATGATGTCCTGCGATGCTCGGCTCTGT	418
Db	419	CTGCTCTGGCTGGATTTTTCAGAGATGGCAATGATGTCCTGCGATGCTCGGCTCTGT	478
QY	419	ATTGTGATTAAGTGTCTTGTCGATGCTACTCTGCGATGATTCCTTCAGACACTTTC	478
Db	479	ATTGTGATTAAGTGTCTTGTCGATGCTACTCTGCGATGATTCCTTCAGACACTTTC	538
QY	479	AGCAGCATCACTGCAACAGCA	501
Db	539	AGCAGCATCACTGCAACAGCA	598
QY	502	-----GAAG	505
Db	599	TGCTCTTTGAGACCTTTGGGGGAAATGGAAGACAAATTTCCAGTGCAGACTGTGAG	658
QY	506	GAGGAGCGTGTG--AAGTGAATAGCAGCACCCGATGTGTAAACAAGATCGCATTGAGAG	564
Db	659	GAGGAGCGTGTGAAGATGATAGCAGCACCGATGTGTAAACAAGATCGCATTGAGAG	718
QY	565	CGGTACAAACAGTAAAGTGTCTGTCTAAGTCTGAAAAGTGGCTGAAACAAGAAG	624
Db	719	CGGTACAAACAGTAAAGTGTCTGTCTAAGTCTGAAAAGTGGCTGAAACAAGAAG	778
QY	625	CGGCTCTTCTGGCGTGCATGCTCATATGATTTGGGAAATGGTGGTGGAGATGAGCCT	684
Db	779	CGGCTCTTCTGGCGTGCATGCTCATATGATTTGGGAAATGGTGGTGGAGATGAGCCT	838
QY	685	TGCTTAAGAGAGAGATGTAAGACACTCCCTGACAAATTCGATGGATGTGGCAACA	744
Db	839	TGCTTAAGAGAGAGATGTAAGACACTCCCTGACAAATTCGATGGATGTGGCAACA	898
QY	745	GGCAACAAAATTAAGACACAGAGATTCACCCAGAACCTTAAAGAGACTTTGTGTAG	804
Db	899	GGCAACAAAATTAAGACACAGAGATTCACCCAGAACCTTAAAGAGACTTTGTGTAG	958
QY	805	TAAGGAA 812	
Db	959	TAAGGAA 966	
RESULT 3			
ID	ABX71301	standard; cDNA; 1936 BP.	
AC	ABX71301;		

XX 14-APR-2003 (first entry)
DT Human kidney-derived cDNA from clone DNFzphfk42_4b6.
XX
DE Human; gene, gene therapy; vaccine; disease treatment; detection; ss.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
PM
PD 22-FEB-2001.
PP
PX 18-AUG-2000; 2000WO-IB001496.
PY
PR 18-AUG-1999; 99US-014949P.
PS 28-SEP-1999; 99US-0156503P.
PT
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-32784C/34.
DX P-PsDB; ABUS2838.
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues; useful in recombinant DNA methodologies.
XX
PS Claim 1; Page 425; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence encodes a polypeptide
CC described in the disclosure of the invention

Sequence 1936 BP; 561 A; 371 C; 408 G; 576 T; 0 U; 0 Other;

Db 726 ACTCCCTGACATTCCTGATGATGATGGGCAACAGCAACAAATTAAGCACAGAGAT 785
 QY 771 TCACCCAGAACTTAACAGAGCATTTGTGTAGTAAGGAA 812
 Db 786 TCACCCAGAACTTAACAGAGCATTTGTGTAGTAAGGAA 827

RESULT 4
 AAKS1800 ID AAKS1800 standard; cDNA; 827 BP.
 AC AAKS1800;
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 345.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 XX
 FN WO200157190-A2.
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00653561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Gao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PsDB; AAM78667.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 1; Page 1351-1352; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAKS1456-AAKS3435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAKS2581), 2111
 CC (AAKS2582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SO Sequence 827 BP; 230 A; 194 C; 198 G; 199 T; 0 U; 6 Other;
 XX
 Query Match 53.0%; Score 430.2; DB 4; Length 827;
 Best Local Similarity 99.3%; Pred. No. 5,4e-117;
 Matches 432; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 378 TCAGAGAAATGGCAATGCTCTGCGATGCTGGGTCCTGTATTGTGATTAAGTCTTG 437
 Db 210 TTATGAATGGCAATGCTCTGCGATGCTGGGTCCTGTATTGTGATTAAGTCTTG 269
 QY 438 TCAGATGCTACTTCTGCTGATGATCCCTTGACACACTTTCACAGAGCATCACTTGACAG 497
 Db 270 TCAGATGCTACTTCTGCTGATGATCCCTTGACACACTTTCACAGAGCATCACTTGACAG 329
 QY 498 ACCAGAGAGGAGGAGTGTGAGTATGAGCAACCGATGTTGTAACAAGATCGAT 557
 Db 330 ACCAGAGAGGAGGAGTGTGAGTATGAGCAACCGATGTTGTAACAAGATCGAT 389
 QY 558 TGAGAGCGGTCACAAACAGTAAGTGTCTGTCTACCTGGAAAAAGTGGTGAACAAC 617
 Db 390 TGAGAGCGGTCACAAACAGTAAGTGTCTGTCTACCTGGAAAAAGTGGTGAACAAC 449
 QY 618 AAGAAAACGGCTCTTCTGCTGATGCTCCATGATGATGGGAAATGGTGTGAGAT 677
 Db 450 AAGAAAACGGCTCTTCTGCTGATGCTCCATGATGATGGGAAATGGTGTGAGAT 509
 QY 678 GAGGCTTGGCTTGAAGAGGAAAGATGTPAAGCACTCCCTGACATTTCTGATGATG 737
 Db 510 GAGGCTTGGCTTGAAGAGGAAAGATGTPAAGCACTCCCTGACATTTCTGATGATG 569
 QY 738 CGCAAGAGGCAACAAATTAAGACAGAGATTCACCCAGAACCTAACAGAGCATTT 797
 Db 570 CGCAAGAGGCAACAAATTAAGACAGAGATTCACCCAGAACCTAACAGAGCATTT 629
 QY 798 GTGTAGTAAGGAA 812
 Db 630 GTGTAGTAAGGAA 644

RESULT 5
 AAKS2784 ID AAKS2784 standard; cDNA; 671 BP.
 AC AAKS2784;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 2313.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 XX
 FN WO200157190-A2.
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00653561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Gao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PsDB; AAM79651.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.

XX Claim 1, Page 4638, 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibitin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX Sequence 671 BP, 182 A, 162 C, 167 G, 160 T, 0 U, 0 Other;

XX Query Match 52.4%; Score 425.4; DB 4; Length 671;
 XX Best Local Similarity 98.6%; Pred. No. 1.3e-115;
 XX Matches 429; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 378 TCAGGAATGGCAATGCTCTGCGCATGCTCGGCTCTGATTTGTGATTAAGCTTG 437
 DB 210 TTATGGAATGGCAATGCTCTGCGCATGCTCGGCTCTGATTTGTGATTAAGCTTG 269
 QY 438 TGCATGCTACTCTGCGCATGATCCCTTCAGACACTTTCAGACATCACTGCACAG 497
 DB 270 TGCATGCTACTCTGCGCATGATCCCTTCAGACACTTTCAGACATCACTGCACAG 329
 QY 498 ACCAGAGAGGAGCGTGTGAAGTATAGCAGACACCCGATGTTTAACAGAAATCGAT 557
 DB 330 ACCAGAGAGGAGCGTGTGAAGTATAGCAGACACCCGATGTTTAACAGAAATCGAT 389
 QY 558 TGAGGAGCGGTCAACAACAGTAAAGTTCCTGCTTACTGTAAGAAAGTGGTGGAAACAC 617
 DB 390 TGAGGAGCGGTCAACAACAGTAAAGTTCCTGCTTACTGTAAGAAAGTGGTGGAAACAC 449
 QY 618 AAGAAACCGGCTCTTCTGGTGCATGCCCTCCATAGATTTGGGAAATGGTGGTGAAT 677
 DB 450 AAGAAACCGGCTCTTCTGGTGCATGCCCTCCATAGATTTGGGAAATGGTGGTGAAT 509
 QY 678 GGAAGCTTGGCTTGAAGAGAGAGATTAAGACACTCCCTGACAAATTCGATGATGTG 737
 DB 510 GGAAGCTTGGCTTGAAGAGAGATTAAGACACTCCCTGACAAATTCGATGATGTG 569
 QY 738 CGCAACAGCAACAATTAAGACCAAGAAATTCACCAAGAACTTAACAGAGATT 797
 DB 570 CGCAACAGCAACAATTAAGACCAAGAAATTCACCAAGAACTTAACAGAGATT 629
 QY 798 GTGGTAGTAAGGAA 812
 DB 630 GTGGTAGTAAGGAA 644

RESULT 6
 AAC08673
 ID AAC08673 standard; cDNA; 261 BP.

XX AAC08673;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 12748.

XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EPI033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000, 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GENSET).

XX Dumas Malne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 12748; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors

XX Sequence 261 BP, 70 A, 72 C, 52 G, 65 T, 0 U, 2 Other;

XX Query Match 23.0%; Score 187; DB 3; Length 261;
 XX Best Local Similarity 98.9%; Pred. No. 4.7e-45;
 XX Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACTGAGATGGGAGTGTGCATGCGCACTATTAATGATTCATAGAACTGACA 60
 DB 73 CACTGAGATGGGAGTGTGCATGCGCACTATTAATGATTCATAGAACTGACA 132
 QY 61 TTATCTCCCATCACTTCAAGAGTCTGTGACGAGAGGAGGCCGAGAGATGATTA 120
 DB 133 TTATCTCCCATCACTTCAAGAGTCTGTGACGAGAGGAGGCCGAGAGATGATTA 192
 QY 121 AGGTGAATAAGCAAGGTTTCAACCTCAACCTTGGCTCTTTTCTGCAATACAGTC 180
 DB 193 AGGTGAATAAGCAAGGTTTCAACCTCAACCTTGGCTCTTTTCTGCAATACAGTC 252
 QY 181 TGAATGAAC 189
 DB 253 TGAATGAAC 261

RESULT 7
 AA158347
 ID AA158347 standard; cDNA; 716 BP.

XX AA158347;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 550.

XX Human; noctropic; immunosuppressant; cyrostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX	Chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM	leukaemia; ss.
OS	Homo sapiens.
XX	
PN	MO200153312-A1.
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US034263.
XX	
PR	23-DEC-1999; 99US-00471275.
PR	21-JAN-2000; 2000US-00468725.
PR	23-APR-2000; 2000US-00552317.
PR	20-JUN-2000; 2000US-00598042.
PR	19-JUL-2000; 2000US-00620312.
PR	03-AUG-2000; 2000US-00653450.
PR	14-SEP-2000; 2000US-00662191.
PR	19-OCT-2000; 2000US-00693036.
PR	29-NOV-2000; 2000US-00727344.
XX	
PA	(HYSEQ-) HYSEQ INC.
PI	Tang Y ^T , Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI	Zhou P, Goodrich R, Drymacac RT;
XX	
DR	WPI; 2001-442253/47.
XX	P-PsDB; AAM39191.
PT	
PT	Novel nucleic acids and polypeptides, useful for treating disorders such
PT	as central nervous system injuries.
PS	Claim 1; SEQ ID NO 550; 10078p; English.
XX	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and the
CC	encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders. Note: The sequence data for this patent did not form
CC	part of the printed specification
XX	
SEQ	Sequence 716 BF; 212 A; 156 C; 195 G; 153 T; 0 U; 0 Other;
XX	
Query Match	16.2%; Score 131.2; DB 4; Length 716;
Best Local Similarity	63.1%; Pred. No. 2.8e-28;
Matches 202; Conservative	0; Mismatches 118; Indels 0; Gaps 0
479	AGCAGATACCTCGCACAGACCGAAGGAGGCGGTGAGATGATAGCAGACACCGAT 538
DB	
295	ACCATATATAAGCTCACCATGTTTAAACGGGAATTGTGAGTGGCACTCCACAGAT 354
479	
539	GTTGTACAGAGATCGCATTTAGAGAGGGGTCAAAACAGTAAAGTCTTCGTCTACTG 598
DB	
355	GCTGTATTAAGAACAGATGAGAAACGGTCAACAAACAGTCAAGTGTCTGCTTCCCTG 414
479	
599	GAAGTGTGCTGGAACAAACAAAGAACCGGCTTCTTGCTGATGCCCTCATAGTGAATG 658
DB	
415	GGAAGTGGAGGCAACACCGCAGTGTCTCCATCATGTGTGATGCTTCAATAGTGAAC 474
479	
659	GGAAATGGTGTGTGATGATGAGCTTGTGCTTGAAGAGAAAGATGTAAACACTCCCTG 718
DB	
475	AGAAATGGTGTGTGATGATGAGCTTGTGCTTGAAGAGAAAGATGTAAAGTCTTCCGG 534
475	

OY		719	ACATTGTTGGATGTGAGTGGCAAAACGGAACAATAATTAGACACAGAATTCACCCAA	778
Db		535	ATCGGAAAAGGATGGAGCTGTTCTCTGGGATTAAGTCATAAACAATCAGGTAAACCATT	594
OY		779	GAACTTAACAGAACGATTGG	798
Db		595	AACCCAGAGAAATCTACACTG	614
RESULT 8				
ID	ADB48315	ADB48315 standard; cDNA; 716 BP.		
XX	ADBA8315;			
AC	ADBA8315;			
XX				
DT	04-DEC-2003 (first entry)			
XX				
DE	Novel human CDNA SEQ ID NO 225.			
XX				
KM	ss; cancer; neurodegenerative disease; human.			
XX				
OS	Homo sapiens.			
XX				
PN	US2003104529-A1.			
PD				
XX	05-JUN-2003.			
PF				
XX	04-JAN-2002; 2002US-00037270.			
PR	21-JAN-2000; 2000US--00488725.			
FR	25-FEB-2000; 2000US--00552317.			
PR	19-JUL-2000; 2000US--00620312.			
XX				
PA	(ZHOU/) ZHOU P.			
PA	(TANG/) TANG Y T.			
PA	(LIUC/) LIU C.			
PA	(ASUNDI/) ASUNDI V.			
PA	(DRMA/) DRMANAC R T.			
PI	Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;			
XX				
DR	WPI; 2003-678194/64.			
XX				
PT	New polynucleotide, useful for treating diseases e.g., cancer or			
PT	neurodegenerative diseases.			
XX				
PS	Claim 1; SEQ ID NO 225; 9pp; English.			
XX				
CC	The invention relates to a polynucleotide comprising a sequence given in			
CC	the specification, or its mature protein-coding portion, or its			
CC	complement. The polynucleotide is useful for treating diseases e.g.,			
CC	cancer or neurodegenerative diseases and many others listed in the			
CC	specification. The present sequence represents a novel human cDNA. Note:			
CC	The sequence data for this patent did not form part of the printed			
CC	specification or was obtained in electronic format directly from USPTO			
CC	at seqdata.uspto.gov/sequence.html?DocID=20030104529.			
XX				
SQ	Sequence 716 BP; 212 A; 156 C; 195 G; 153 T; 0 U; 0 Other;			
Query Match	16.2%; Score 131.2; DB E; Length 716;			
Best Local Similarity	63.1%; Pred. No. 2.8e-28;			
Matches 202; Conservative	0; Mismatches 118; Indels 0; Gaps 0			
OY	479	AGAGAGTCACTGCACAGACACAGAGAGGAGCGTGTAAGTAGACGACACACCAT	538	
Db	295	ACCATCATTAAGTCACTGCCATGTTAAACCGGAACCTTGAGTGTGGCACATCCACAGAT	354	
OY	539	GTTGTAAACAAGATGCATTGAGAGCGGTCACAACACAGTAAAGTGTCTGTCTTAACTG	598	
Db	355	GGGTGTAATAAGAACAGATPAGAAGAACGGTCAACAACAGTCAAAGTGCTCTCTCCCTG	414	
OY	599	GAAGTGGCTGGACACAACAAGAACCGGCTTTCTTGCTGCTCATGCTCTCATAGCATGG	658	

Db 415 GGCAGGTGGCAGGACACGAGCTGCTCCATCATGTGTGATGTTCAATAGTGAAC 474
 Qy 659 GGAATGTTGTTGTTGATGATGAGCTTCTGCTTACAGAGAAAGATGTAAGACCTCCTG 718
 Db 475 AGAAATGTTGTTGCTCCATGATGACCATGCTTGAAGGAGAAAGATGTAAGTTCTCCGG 534
 Qy 719 ACAATCTGATGATGTTGCGCAACAGCAACAAATTTAGACCGAGAAATTCACCCAA 778
 Db 535 ATCGGAAAGATGAGAGCTGTTCTCTGGAAATTAAGTCAAAACAATGAGGTAACTT 594
 Qy 779 GAACCTAACGAAAGCAATTG 798
 Db 595 AACCCAGAGAAATCAAGTG 614
 RESULT 9
 AAI60133
 ID AAI60133 standard; cDNA; 1171 BP.
 AC AAI60133;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 4122.
 XX
 KM Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-0048725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,
 PI Zhou F, Goodrich R, Drmanac RT;
 XX
 WP1: 2001-444253/47.
 P-PSDB; AAM40977.
 XX
 FT Novel nucleic acids and polypeptides, useful for treating disorders such
 FT as central nervous system injuries.
 XX
 PS Claim 1; SEQ ID NO 4122; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SO Sequence 1171 BP; 329 A; 265 C; 321 G; 254 T; 0 U; 2 Other;
 Query Match 16.2%; Score 131.2; DB 4; Length 1171;
 Best Local Similarity 63.1%; Pred. No. 3.5e-28;
 Matches 202; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
 Qy 479 AGCAGCATCCTGACAGACAGGAGAGGAGCGTGTGATGATGACAGACGAGT 538
 Db 673 ACCATATTAAGCTCACCATGTTAAAGGAACTGTGAGTGTGACCTCACAGAT 732
 Qy 539 GTTGTAAAGAAATGCAATTTGAGAGCGGTCAAAACAGTAAGTTCTGTTACCTG 598
 Db 733 GCTGTATTAAGAAAGAAATGAAAGAAAGCGTCAAAACAGTCAAGTCTCTCCCTG 792
 Qy 599 GAAAAGTGTGCTGACACACAAAGAAACCGGCTTTCTGTGTTGATGCTCCATGATG 658
 Db 793 GGCAGGTGGCAGGACACAGCGAGCTGCTCCATCATGTTGTGATGCTTCAATGATG 852
 Qy 659 GGAATGTTGTTGTTGATGATGAGCTTGCCTTGAAGAGAGAAATGTAAGACACTCCCTG 718
 Db 853 AGAAATGTTGTTGTTGCTCCATGATGAGCTATGCTTGAAGAGAAAGATGTAAGTTCTCCGG 912
 Qy 719 ACAATCTGATGATGATGATGCGGCAACAGCAACAAATTTAGACCGAGAAATTCACCCAA 778
 Db 913 ATCGAAGAGATGAGAGCTGTTCTCTGGAATTAAGTCAAAACAATGAGGTAACTT 972
 Qy 779 GAACCTAACGAAAGCAATTG 798
 Db 973 AACCCAGAGAAATCAAGTG 992
 RESULT 10
 ADB62738
 ID ADB62738 standard; cDNA; 2390 BP.
 XX
 AC ADB62738;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA encoding clone NT2NE20069580.
 XX
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 PI Key Location/Qualifiers
 PI CDS 1404..2282
 FT /*tag= a
 FT /product= "Clone NT2NE20069580 protein"
 XX
 EP1308459-A2.
 XX
 PD 07-MAY-2003.
 XX
 PE 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX WPI; 2003-450961/43.
DR P-PSDB; ADB64708.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX

Claim 1, Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC medicines may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX

Sequence 2390 BP; 524 A; 629 C; 527 G; 710 T; 0 U; 0 Other;

Query Match 15.8%; Score 128.4; DB 9; Length 2390;
Best Local Similarity 92.5%; Pred. No. 3.3e-27;

Matches 135; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY	357	CTCTGCTCTTGAGCTGATTTTTCAGAGATGCAATGGTCTCTGCAATGCTGGTCTCT	416
DB	610	CTTTCCTGCTCAATGTTCTCTTTAGAGATGCAATGGTCTCTGCAATGCTGGTCTCT	669
QY	417	GTAATTTGATTAAGTCTGTTGCAATGCTACTGTCATGCAATGCTGCTTCAACACTTT	476
DB	670	GTAATTTGATTAAGTCTGTTGCAATGCTACTGTCATGCAATGCTGCTTCAACACTTT	729
QY	477	CCAGCAGCATCACCTGCACAGCCAG 502	
DB	730	CCAGCAGCATCACCTGCACAGCCAG 755	

RESULT 11

AAC25659
ID AAC25659 standard; cDNA; 238 BP.

XX AAC25659;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 29734.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP103401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GIST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

Claim 1, SEQ ID NO 29734; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX

Sequence 238 BP; 69 A; 47 C; 72 G; 50 T; 0 U; 0 Other;

Query Match 13.5%; Score 109.6; DB 3; Length 238;
Best Local Similarity 69.8%; Pred. No. 4.2e-22;

Matches 162; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY	503	AAGAGGAGCTGTGTAAGTATGATGAGCAGCACCAGATGTTGTAAACAAGATGCAATTGAGG	562
DB	2	AACGGGAATCTGTAGAGTGTGAGCAGCAGTCCAGATGCTGTAAAGAAACAAGATGAGG	61
QY	563	AACGGTCAACAAACAGTAAGTGTCTCTGCTTACTGCAAAAAGTGGCTGCAACAAAGAA	622
DB	62	AACGGTCAACAAACAGTAAAGTGTCTCTGCTTACTGCAAAAAGTGGCTGCAACAAAGAA	121
QY	623	ACCG-ACCTTCTTGGCTGATGCTCCATAGATGGAATGGTGGTGGATGGAG 681	
DB	122	GCTGCTCCATCATAGTGTGATGCTCTCAATAGTGAACAGAAATGGTGGTGGATGGAG	181
QY	682	CCTTGCTTGAAGAGGAAGATGTAAGACACTCCCTGACATCTGCTGATGA 733	
DB	182	CCATGCTTGAAGAGGAAGATGTAAGTCTTCCGATCCGAAAGATGGA 233	

RESULT 12

AAX51967
ID AAX51967 standard; DNA; 333 BP.

XX AAX51967;

XX 22-JUN-1999 (first entry)

DE Human secreted protein 5' EST SEQ ID NO: 181.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KM forensic; gene therapy; chromosome mapping; signal peptide;
KM upstream regulatory sequence; cytokine activity; cell proliferation;
KM differentiation; haematopoiesis regulation; tissue growth regulation;
KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KM thrombolytic; anti-inflammatory; tumour inhibition; ds.

XX Homo sapiens.

XX


```
OY 508 GGGACGTGTGAGTATGACAGACACCCGATGTTGTACAGAAATCCGATTGAGAGCGG 567
Db 121 GGCACCTGTGTGAGATTGTGACCTTGGACCGGACAGACAGCCCTCGAGGACGATCGCC 180
OY 568 TCACAAACAGTAAAGTGTTCCTGTCTACCTGSAAAAAGTGCTGSAACAACAAGAAACCGG 627
Db 181 CGGACAGCCGCCCGCTGTGCGTGTGAAAAGGGGCGAGATCGCCGGACCCACAGAGCCCGG 240
OY 628 CTTTCTGTGCGTGTGATGCTCCATAGTGAATGGGAAAATGATGTGTGAGATGAGCTTGC 687
Db 241 CCGGCTGTGTGAGAGCAAGATCATCAAGACCAAGCAAGTGTGTGACATGCTTCCGTGT 300
OY 688 CTAGAAGGAGAAAGAAATGTAAAGCACTCCCTGACAAATTTCTGANTGATGTGCGC 740
Db 301 CTGAGAGGGGGAAGCTGCGACTTGTTAATCAACCGGTCAAGCTGAGCGTGCAC 353
```

Search completed: April 9, 2004, 15:02:20
Job time : 405 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 09:39:26 ; Search time 391 Seconds

(without alignments)
10634.179 Million cell updates/sec

Title: US-09-763-335-4

Perfect score: 812

Sequence: 1 cacttgagtcg99gagtcgtcc.....catttgtagtaagaa 812

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Genbank1:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_ov:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rsd:*

36: em_htg_mam:*

37: em_htg_vtl:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	450.2	55.4	801	10	AY325120	AY325120 Mus muscu
2	433.2	53.3	1936	6	AX086349	AX086349 Sequence
3	433.2	53.3	1936	6	HSMB03068	AL713702 Homo sapi
4	431.8	53.2	815	9	AY325114	AY325114 Homo sapi
5	368.8	45.4	137703	9	AC092036	AC092036 Homo sapi
6	368.8	45.4	158080	9	AC138390	AC138390 Homo sapi
7	368.8	45.4	187748	9	AC139766	AC139766 Homo sapi
8	215.2	26.5	247515	2	AC116080	AC116080 Rattus no
9	190.2	23.4	193027	10	AC117234	AC117234 Mus muscu
10	187	23.0	261	6	AX896885	AX896885 Sequence
11	187	23.0	261	6	BD032418	BD032418 Sequence
12	145	17.9	173288	9	AC104167	AC104167 Homo sapi
13	145	17.9	173288	9	AC112214	AC112214 Homo sapi
14	137.4	16.9	901	10	AY325123	AY325123 Mus muscu
15	136.8	16.8	1897	9	AK057890	AK057890 Homo sapi
16	136.8	16.8	2188	9	BC031566	BC031566 Homo sapi
17	136.8	16.8	2274	9	AY325117	AY325117 Homo sapi
18	135.8	16.7	801	10	AY325122	AY325122 Mus muscu
19	134.2	16.5	2011	9	AB063062	AB063062 Macaca fa
20	134.2	16.5	2265	9	AB045997	AB045997 Macaca fa
21	132.6	16.3	1077	9	AY325116	AY325116 Homo sapi
22	131.2	16.2	671	9	AY325115	AY325115 Homo sapi
23	131.2	16.2	716	6	AR338734	AR338734 Sequence
24	131.2	16.2	2718	9	AK123580	AK123580 Homo sapi
25	131.2	16.2	3474	9	HSMB05409	AL834160 Homo sapi
26	131.2	16.2	4058	9	BC050347	BC050347 Homo sapi
27	128.4	15.8	2390	6	AX747367	AX747367 Sequence
28	128.4	15.8	2390	9	AK092129	AK092129 Homo sapi
29	128	15.8	3087	9	BC028403	BC028403 Homo sapi
30	127.6	15.7	43205	9	AC126123	AC126123 Homo sapi
31	124	15.3	1101	10	AY325121	AY325121 Mus muscu
32	124	15.3	4039	10	BC027082	BC027082 Mus muscu
33	122.6	15.1	243881	2	AC135662	AC135662 Rattus no
34	121.2	14.9	207096	2	AC102744	AC102744 Mus muscu
35	109.6	13.5	238	6	AX913871	AX913871 Sequence
36	109.6	13.5	238	6	BD049404	BD049404 Sequence
37	100.8	12.4	158186	2	BX890603	BX890603 Danto rer
38	88.8	10.9	333	6	BD077927	BD077927 5'EST of
39	88.2	10.9	179979	2	BX571952	BX571952 Danto rer
40	80.2	9.9	137393	2	BX545852	BX545852 Danto rer
41	80.2	9.9	163795	2	BX545818	BX545818 Danto rer
42	79.4	9.8	218896	2	AC106975	AC106975 Rattus no
43	77.2	9.5	171705	9	AC104166	AC104166 Homo sapi
44	77.2	9.5	218652	9	AC096922	AC096922 Homo sapi
45	76.8	9.5	212238	2	AC134077	AC134077 Rattus no

ALIGNMENTS

RESULT 1

AY325120

LOCUS AY325120 801 bp mRNA linear ROD 23-JUL-2003

DEFINITION Mus musculus TAFAl mRNA, complete cds.

ACCESSION AY325120

VERSION AY325120.1 GI:32967238

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

REFERENCE

1 (bases 1 to 801)

Tang,Y.T., Emtage,P., Funk,W., Hu,T., Atteburn,M., Park,E. and Rupp,F.

TAFAl: A Novel Secreted Family with Homology to CC-chemokines

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 801)
AUTHORS Tang, Y.T., Emrige, P., Funk, W., Hu, T., Arterburn, M., Park, E. and Rupp, F.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Biology Research, Nuvolo, 675 Almanor Avenue, Sunnyvale, CA 94085, USA
FEATURES
source location/Qualifiers
1..801
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
238..639
/note="Secreted protein similar to CC-chemokines"
/codon_start=1
/product="TAF1"
/protein_id="AAP2412.1"
/db_xref="GI:32967239"
/translation="MAMVSAMSRLYIMISACMLCHSLQHTFOGHLHREGGTC
EVIARHCNKNRIERSQVKSCLPEKVAGITRNPSCVDASIVIGKWCMEPOL
EGEBCKTLPDNSGWCATGNKIKTRHPRT"

ORIGIN

Query Match 55.4%; Score 450.2; DB 10; Length 801;
Best Local Similarity 86.4%; Pred. No. 8.5e-109;
Matches 497; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 238 TTTTATTTTATCTGATTAAGAAGATTGTGGGAAAGCTTTGAAAAATTTTAA 297
DB 91 TTTTATTTTCAATCTTAATAAAGATTGCTGGGAAGCATTTCTTGAATAATCTGAA 150
QY 298 TTGTGGCAGATGATTTTAAAAAGTTAGATCTTTCAATGACATTAAGAGTAC 357
DB 151 CTGTGGCAGATGATTTTAAAAAGTTAGATCTTTCAATGACATGAGAGGATTC 210
QY 358 TGTGCTCTTGCTGATTTTCAAGATGGAATGCTCTGCGATGCTCTGCTGCTG 417
DB 211 CCGTCTCTTGCTGATTTTCAAGATGGAATGCTCTGCGATGCTCTGCTGCTG 270
QY 418 TATTGTGGATTAAGCTTTGCAATGCTACTGCTGCTGCTGCTGCTGCTGCTGCT 477
DB 271 TACTGTGGATTAAGCTTTGCAATGCTACTGCTGCTGCTGCTGCTGCTGCTGCT 330
QY 478 CAGCAGCATCACTGCAAGACCAAGAGAGGAGAGCTGTGAATGATGACAGCAGCA 537
DB 331 CAGCAGCATCACTGCAAGACCAAGAGAGGAGAGCTGTGAATGATGAGGAGGAG 390
QY 538 TGTGTGAACAGATGCTGATGAGAGGCTGCAAAAGTAAAGTGTCTCTGCTACCT 597
DB 391 TGTGTGAACAGATGCTGATGAGAGGCTGCAAAAGTAAAGTGTCTCTGCTACCT 450
QY 598 GGAAGAGGCTGCAAGACCAAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCT 657
DB 451 GGAAGAGGCTGCAAGACCAAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCT 510
QY 658 GGAAGAGGCTGCTGATGAGAGGCTTGCCTGAGAGAGAGAGAGAGAGAGAGAG 717
DB 511 GGAAGAGGCTGCTGATGAGAGGCTTGCCTGAGAGAGAGAGAGAGAGAGAGAG 570
QY 718 GACAAATTCGATGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777
DB 571 GACAAATTCGATGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
QY 778 AGAAGCTTAACAG 812
DB 631 AGAAGCTTAACAG 665

RESULT 2
AX086349 1936 bp DNA linear PAT 09-MAR-2001
LOCUS AX086349
DEFINITION Sequence 301 from Patent WO0112659.

ACCESSION AX086349
VERSION AX086349.1 GI:13275945
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Wiemann, S.
TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 301 22-FEB-2001;
German Human Genome Project (HG)
FEATURES
source location/Qualifiers
1..1936
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 53.3%; Score 433.2; DB 6; Length 1936;
Best Local Similarity 96.1%; Pred. No. 2.9e-104;
Matches 444; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 351 AGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 410
DB 366 AGGTAACTCTTCTCTGCAATGTTCTTTAGAGATGCAATGCTCTGCAATGCTCTG 425
QY 411 GGTCTGATTTGTGATTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
DB 426 GGTCTGATTTGTGATTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
QY 471 CACTTTCAGCAGCATCACTGCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAG 530
DB 486 CACTTTCAGCAGCATCACTGCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAG 545
QY 531 ACACGAGTGTGAACAGATGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 590
DB 546 ACACGAGTGTGAACAGATGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
QY 591 TCTACCTGAAAAAGTGTGCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 650
DB 606 TCTACCTGAAAAAGTGTGCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 665
QY 651 AGTATGAGAGAGAGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 710
DB 666 AGTATGAGAGAGAGTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
QY 711 ACTCCCTGCAATTTCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 770
DB 726 ACTCCCTGCAATTTCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
QY 771 TCACCCAG 812
DB 786 TCACCCAG 827

RESULT 3
HSM803068 1936 bp mRNA linear PRI 20-MAR-2002
LOCUS HSM803068
DEFINITION Homo sapiens mRNA, cDNA DKFP556B064 (from clone DKFP556B064).
ACCESSION AL713702
VERSION AL713702.1 GI:19584419
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1936)
AUTHORS Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPs, Am Kioferepitz 18a, D-82152 Martinsried, GERMANY

Ox 798 GTGCTAGTAAGGAA 812
 Db 630 GTGCTAGTAAGGAA 644

RESULT 5
 AC092036/c 137703 bp DNA linear PRI 13-SEP-2001

LOCUS Homo sapiens chromosome 3 clone RP11-79C12, complete sequence.

AC092036 AC0923496

AC092036.3 GI:15594336

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 137703) Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,

Clendenning, V., and Haugen, E.D.

Unpublished

2 (bases 1 to 137703)

Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and

Haugen, E.D.

Direct Submission

Submitted (16-JUN-2001) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 137703) Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,

Clendenning, V., and Haugen, E.D.

Direct Submission

Submitted (13-SEP-2001) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

On Sep 13, 2001 this sequence version replaced gi:15138904.

----- Genome Center

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Contact: uwgcts@u.washington.edu

Drafting Center: BCM

----- Project Information

Center project name: chr-3

Center clone name: RP11-79C12 (bc0194)

----- Summary Statistics

Sequencing vector: M13; L08821; 60% of reads

Sequencing vector: plasmid; L08752; 40% of reads

Chemistry: Dye-terminator Big Dye; 28% of reads

Chemistry: Dye-terminator Big Dye; 28% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 139753 bases at least Q40

Consensus quality: 140133 bases at least Q30

Consensus quality: 140304 bases at least Q20

Insert size: 137703; sum-of-contigs

Quality coverage: 9.6x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5': RP11-81N13 (UWGC:bc0203) AC016935

3': RP11-364N10 (UWGC:bc0385) AC021437

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII
 EcoRI
 HindIII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
2273	2272	8696	8904	2190	2099
2067	2102	6	<800	6382	6589
6314	6524	159	<800	512	<800
813	778	1461	1466	449	<800
9138	8978	839	848	5669	5435
4117	4135	1014	994	150	<800
5429	5600	458	<800	680	<800
3261	3286	1505	1486	457	<800
11041	10746	968	994	3181	3182
9336	9295	55	<800	3742	3733
92	<800	2448	2471	1207	1241
6268	6269	1679	1736	2016	1899
5523	5600	6963	6941	56	<800
17081	17253	8613	8629	2083	2099
509	<800	4027	3975	3015	3182
768	778	6850	6941	3204	3306
1815	1827	2003	2026	5520	5435
2712	2727	9934	9883	536	<800
5619	5600	1410	1486	2744	2830
1410	1393	1784	1736	460	<800
6945	7000	1225	1249	3609	3588
1947	1947	1719	1736	2707	2830
286	<800	698	<800	10577	10789
2123	2102	3448	3300	892	879
501	<800	3219	3300	5440	5435
5108	5108	1739	1736	6194	6078

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred

Overlapping Sequences:

5': RP11-79C12 (UMGC:bc0194) AC092036, 134230-bp overlap
3': RP11-89A12 (UMGC:bc0207) AC092416, 45454-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI

HindIII

BglII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8768	7698	7651	5136	5080
6	<800	6382	6470	2067	2084
458	<800	512	<800	12792	12692
1505	1524	449	<800	4117	4117
968	946	2196	2138	5429	5542
55	<800	150	<800	3261	3294
2448	2480	680	<800	11041	10803
1679	1712	457	<800	9336	9287
6963	6907	3181	3188	92	<800
8613	8768	3742	3737	6268	6223
4027	3995	1207	1229	5523	5542
6850	6907	2016	1912	17081	17289
2003	2014	56	<800	509	<800
9934	9693	2083	2138	768	779
1410	1413	3015	3023	1815	1794
1784	1712	3204	3188	2712	2734
1225	1225	5520	5367	5619	5542

1719	1712	536	<800	1410	1367
698	<800	2744	2824	6945	6947
3448	3361	460	<800	1947	1919
3219	3361	3609	3585	286	<800
1739	1712	2707	2824	2123	2084
4504	4516	10577	10818	501	<800
715	<800	892	893	5108	5080
1791	1712	5440	5367	3326	3294
3659	3686	6194	6047	5600	5542
893	946	1504	1447	116	<800
2787	2820	873	893	2897	2936
1973	2014	1299	1229	5688	5542
7554	7429	3318	3316	771	779
4961	4996	2913	2824	2951	2936
281	<800	423	<800	2714	2734
5348	5258	113	<800	91	<800
7332	7429	117	<800	4860	4824
5781	5807	426	<800	4204	4117
1691	1712	1913	1912	588	<800
2611	2663	5403	5367	7730	7706
419	<800	42	<800	2853	2936
3343	3361	9136	9077	1229	1203
443	<800	556	<800	6500	6512
9705	9693	2154	2138	1218	1203
1548	1524	12	<800	1102	1093
5322	5258	4489	4377	81	<800
818	828	2882	2824	274	<800
1425	1413	677	<800		
6636	6599	217	<800		
714	<800	4592	4557		
6726	6907	464	<800		
1243	1225	799	<800		
3504	3361	204	<800		
3363	3361	885	893		
242	<800	585	<800		
		2100	2138		
		4115	4117		

Query Match	45.4%;	Score 368.8;	DB 9;	Length 158080;
Best Local Similarity	98.7%;	Pred. No. 4.2e-87;		
Matches 383;	Conservative 0;	Mismatches 2;	Indels 3;	Gaps 1.

[illegible]

Location/Qualifiers
1. .187748

```

repeat_region      /organelle=Homo sapiens
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="X3"
                    /clone="RP11-688P9"
                    52..82
repeat_region      /rpt_family="(TA)n"
                    complement(83..379)
                    /rpt_family="Alusq"
                    467..691
repeat_region      /rpt_family="L2"
                    1206..1312
repeat_region      /rpt_family="Tigere6"
                    1985..2012
repeat_region      /rpt_family="AT_rich"
                    2015..2744
repeat_region      /rpt_family="LIMA3"
                    2745..2873
repeat_region      /rpt_family="MER85"
                    2874..4418
repeat_region      /rpt_family="L1MA3"
                    4422..5898
repeat_region      /rpt_family="L1P8"
                    5904..6359
repeat_region      /rpt_family="L1MA3"
                    6332..6397
repeat_region      /rpt_family="AT_rich"
                    7091..7518
repeat_region      /rpt_family="L1ME3B"
                    7576..7597
repeat_region      /rpt_family="(TC)n"
                    7598..7702
repeat_region      /rpt_family="(TA)n"
                    7892..7938
repeat_region      /rpt_family="AT_rich"
                    8102..8124
repeat_region      /rpt_family="(CAAA)n"
                    complement(8554..8981)
repeat_region      /rpt_family="MLT/C"
                    complement(8992..9346)
repeat_region      /rpt_family="TRETA"
                    9385..9900
repeat_region      /rpt_family="L1ME3B"
                    9906..10003
repeat_region      /rpt_family="L1ME3B"
                    10148..10173
repeat_region      /rpt_family="AT_rich"
                    10255..10489
repeat_region      /rpt_family="MIR"
                    10820..10853
repeat_region      /rpt_family="AT_rich"
                    complement(11238..11419)
repeat_region      /rpt_family="MRB"
                    12049..12370
repeat_region      /rpt_family="MIR"
                    12516..12615
repeat_region      /rpt_family="MIR"
                    13267..13439
repeat_region      /rpt_family="L2"
                    13479..13509
repeat_region      /rpt_family="(CAA)n"
                    13539..13620
repeat_region      /rpt_family="MER20"
                    13621..13889
repeat_region      /rpt_family="L1MA8"
                    complement(14577..14883)

```

Matches 383; conservative 0; mismatches 2; indels 3; gaps 1.

Db 48573 TCTGCTCTGGCTGATTTCAGTAA 48600

DEFINITION Rattus norvegicus clone CH230-40C18, *** SEQUENCING IN PROGRESS

ORGANISM *Rattus norvegicus*

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Barnham, F., Bawato, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Detamo, C., Ding, Y., Dinh, H., Diya, K., Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., Fraser, C.M., Gabisi, A., Gante, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunatunne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joliver, A., Kapathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokkelumeh, O., Okwunonu, G., Olampunsaogon, A., Pal, S., Parks, K., Paeternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindeexter, A., Popovic, D., Primus, E., Pu, L.L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, R., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Silter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, C., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 247515)
Submitted (24-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 247515)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 10, 2002 this sequence version replaced g1:21736959.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNWA
Center clone name: CH230-40C18
----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 22934 bases at least Q40
Consensus quality: 23195 bases at least Q30
Consensus quality: 23282 bases at least Q20
Estimated insert size: 236904; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 247515: contig of 247515 bp in length.
Location/Qualifiers
1. 247515
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-40C18"
1. 1162
/note="wgs_end_extension
clone_end:T7"
2343. 4221
/note="wgs_end_extension
clone_end:T7"
5130. 6933
/note="wgs_end_extension
clone_end:T7"
8859. 9540
/note="clone boundary
clone_end:T7
site:ECORI
end_sequence:BM315138"
181526. 183635
/note="wgs contig"
243698. 245700
/note="wgs contig"
245971. 247515
/note="wgs contig"
ORIGIN
Query Match 26.5%; Score 215.2; DB 2; Length 247515;
Best Local Similarity 82.4%; Pred. No. 3,2e-46;
Matches 325; Conservative 0; Mismatches 53; Indels 18; Gaps 6;

OY 1 CACGTGAGTGGGATGGTCATCGGCAACTATTAAGATTCTCATCAGGAATCGACAC 60
DB 141536 CACGTGAGTGGGATGGTCATCGGCGATTAAGTCTCTTCTTCAAGAACT--ACA 141579
OY 61 TTATCTCCCATCATCTTCAAGGTCCTGTCAGGACAGGTGACGACGACGATGATTAA 120
DB 141578 TTAGCTACACATCATCTTCAAGGTCCTGTCAGGACAGGTGACGACGATGATTAA 141519
OY 121 AGGTGAATGACAAAGTTTCCACCCCTCAAACTTGGCTCTTTTCTGACATACAGTC 180
DB 141518 AGGTGAATGACAAAGTTTCCACCCCTCAAACTTGGCTCTTTTCTGACATACAG-- 141461
OY 181 TGATGAACCGGATGCTTTTCTTTTCTGTAATAGATGGAAGAGATTAACATT 240

Dd	141460	TGAACGACCAGCAGTC--TTCCAACTGTGGAAAACAGACTGGAAAGGAAAGCGTGC	141460
Qy	241	TTTTTT-----AATCTGATTAGA--GATTGTGGAAAGCTCTTTGAAAAA	290
Dd	141403	TTTTTTTTCTTCCAAATCTTAATAAGACTGCTGGAAAGCATTTCTTTGAAAAA	141344
Qy	291	TTTTTA-AATGTGGACACAGATGATTTAAAAAGTGATCTTTCCATGAACAACATA	349
Dd	141343	TCACAGACTGTGGACACAGATGATTTAAAAAGTGTATNGCTCTTCCAAATGAACACTAG	141284
Qy	350	TAAAGTACTCTGCTCTTTGCGTGATTTTCAAGAA	385
Dd	141283	GAGAGTGTCCTGCTCTTGCTGGATTTTCAAGTAA	141248
RESULT 9			
AC117234/c	AC117234	193027 bp	DNA linear ROD 25-NOV-2003
LOCUS	Mus musculus BAC clone RP23-410J2 from chromosome 6, complete sequence.		
DEFINITION	AC117234.4	GI:34495048	
ACCESSION			
VERSION			
KEYWORDS	HTG.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 193027)		
AUTHORS	Swearengen-Shahid,S., Kotlowicz,A. and Schatzkemer,K.		
JOURNAL	The sequence of Mus musculus BAC clone RP23-410J2		
TITLE	Unpublished (2001)		
REFERENCE	2 (bases 1 to 193027)		
AUTHORS	Wilson,R.		
JOURNAL	Sequencing of Mus musculus		
TITLE	Unpublished (2001)		
REFERENCE	3 (bases 1 to 193027)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
JOURNAL	Direct Submission		
TITLE	Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forrest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 193027)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
JOURNAL	Direct Submission		
TITLE	Submitted (12-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	5 (bases 1 to 193027)		
AUTHORS	Wilson,R.K.		
JOURNAL	Direct Submission		
TITLE	Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	6 (bases 1 to 193027)		
AUTHORS	Wilson,R.		
JOURNAL	Direct Submission		
TITLE	Submitted (25-MCV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	On Sep 6, 2003 this sequence version replaced gi:21735140.		
COMMENT			

chemistry, or covered by high quality data (i.e., paired quality >= 30), an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION: The RPCT-23 BAC Library has been constructed by Kazutomo Oosagawa and Minko Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone

FEATURES	Location/Qualifiers
source	1. .193027

```

repeat_region
/rpc_family="M1"
1.128
/rpc_family="L1"
129.197
repeat_region
/rpc_family="B4"
528.737
repeat_region
/rpc_family="MER2_type"
959.1031
repeat_region
/rpc_family="ID"
1032.1075
repeat_region
/rpc_family="B4"
3251.3605
repeat_region
/rpc_family="L1"
3640.3834
repeat_region
/rpc_family="L1"
3917.3988
repeat_region
/rpc_family="MER1_type"
4026.4243
repeat_region
/rpc_family="L1"
4708.4969
repeat_region
/rpc_family="RNER15"
4903.5013
repeat_region
/rpc_family="B4"
5933.6355
repeat_region
/rpc_family="MLR"
7471.7535
repeat_region
/rpc_family="MIR"
8988.9064
repeat_region
/rpc_family="ERVK"
9260.9476
repeat_region
/rpc_family="MLR"
9600.9719
repeat_region
/rpc_family="MLR"
11657.11695
repeat_region
/rpc_family="L2"
11697.12155
repeat_region
/rpc_family="L1"
12124.12182
repeat_region
/rpc_family="L1"
12185.12849
repeat_region
/rpc_family="L1"
12860.12895
repeat_region
/rpc_family="L1"

```


Db 253 TGAATGAAC 261

|||||

RESULT 11

BD032418

LOCUS BD032418 261 bp DNA linear PAT 27-AUG-2002

DEFINITION Sequence tag and encoded human protein.

ACCESSION BD032418 GI:22574160

VERSION BD032418.1

KEYWORDS JP 2001269182-A/8664.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.

TITLE Sequence tag and encoded human protein.

JOURNAL Patent: JP 2001269182-A 8664 02-OCT-2001;

COMMENT

OS Homo sapiens (human)

PN JP 2001269182-A/8664

PD 02-OCT-2001

PF 24-FEB-2000 JP 2000118773

PR 26-FEB-1999 US 60/122487

PI JUAN BAPTIST DUMAS MILNE EDWARDS, BIMERIC DUCLAIR, JEAN YVES

PI JORDAN

PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC

C12N5/10,

PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N5/00, C12N5/00, PC

G06F15/40

CC

FC

Key Location/Qualifiers.

1.261

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 23.0%; Score 187; DB 6; Length 261;

Best Local Similarity 98.9%; Pred. No. 9,4e-39;

Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACTGAGAGGGGATGTCATCGGCACTATTAATCTGATCTCATCGAAGAACTGCACA 60

Db 73 CACTGAGAGGGGATGTCATCGGCACTATTAATCTGATCTCATCGAAGAACTGCACA 132

QY 61 TTATCTCCCATCTCACTTCAAGAGTCTCGTCAGAGAGTGAAGCCAGAGATGATTAA 120

Db 133 TTATCTCCCATCTCACTTCAAGAGTCTCGTCAGAGAGTGAAGCCAGAGATGATTAA 192

QY 121 AGGTGAAGTGAAGAGTTTCAACCCCTTAAACCTTGCTCTTTTTCGACATACAGTC 180

Db 193 AGGTGAAGTGAAGAGTTTCAACCCCTTAAACCTTGCTCTTTTTCGACATACAGTC 252

QY 181 TGAATGAAC 189

Db 253 TGAATGAAC 261

RESULT 12

AC104167

LOCUS AC104167 173288 bp DNA linear PRI 07-FEB-2002

DEFINITION Homo sapiens chromosome 3 clone RP11-253K11, complete sequence.

ACCESSION AC104167 AC004915

VERSION AC104167.2

KEYWORDS GI:18583922

SOURCE HTG.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 173288)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

Direct Submission

Unpublished

2 (bases 1 to 173288)

Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.

Submitted (05-DEC-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 173288)

Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

Direct Submission

Submitted (07-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

On Feb 7, 2002 this sequence version replaced g1:17352431.

----- Genome Center

Center: University of Washington Genome Center

Center Code: UMG

Web site: <http://www.genome.washington.edu>

Contact: uwgchgs@u.washington.edu

Drafting Center: BCM

----- Project Information

Center project name: chr-3

Center clone name: RP11-253K11 (bc0635)

----- Summary Statistics

Sequencing vector: unknown; 50% of reads

Sequencing vector: plasmid; 50% of reads

Chemistry: Dye-terminator ET; 80% of reads

Chemistry: Dye-terminator Big Dye; 20% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173003 bases at least Q40

Consensus quality: 173181 bases at least Q30

Consensus quality: 173237 bases at least Q20

Insert size: 173288; sum-of-contigs

Quality coverage: 10.4x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RP11-306N11 (UMGC:bc0642) AC104443

3': RP11-40E22 (UMGC:bc0611) AC046199

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:

All regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

ECORI

HandIII

BglII

SeqDerMap FngFrnt SeqDerMap FngFrnt SeqDerMap FngFrnt

8696	8884	3187	3244	3162	3246
6	<800	6382	6545	2067	2082
1870	1827	512	<800	5399	5310
4820	4754	449	<800	1097	1116
7995	7888	15274	15715	672	<800
142	<800	7177	7163	2783	2823
3847	3775	2682	2895	4088	4076
3019	2984	2876	2895	3330	3246
6844	6834	726	<800	2082	2082
3829	3775	2563	2650	4622	4580
2135	2129	8723	8629	5562	5484
1430	1386	695	<800	9557	9643
115	<800	118	<800	1062	1062
856	888	1076	988	6101	6068
804	836	416	<800	9870	9643
9439	9526	3042	3069	1818	1764
2090	2129	1634	1640	2394	2443
2019	2036	4516	4487	428	<800
2385	2394	803	742	1155	1116
303	<800	787	<800	1902	1922
1188	1213	556	<800	5063	4970
3021	2984	4804	4668	699	<800
314	<800	46	<800	2018	2082
1459	1386	1688	1730	3256	3246
1101	1116	2177	2209	2895	2959
517	<800	3062	3069	609	<800
2166	2129	6219	6140	1121	1116
2429	2394	7221	7163	3560	3591
2440	2538	10686	10536	7604	7678
123	<800	271	<800	6427	6466
109	<800	1694	1730	2443	2443
7704	7888	3465	3506	1053	1062
20	<800	2013	2043	1419	1418
1404	1386	2869	2895	2619	2644
1420	1386	5460	5437	2106	2082

4677	4625	945	883	1128	1116
1141	1116	562	<800	829	850
3303	3278	4205	4172	3107	3246
819	836	315	<800	3283	3246
4124	4126	525	<800	10528	10519
883	888	9661	9490	2163	2249
2305	2394	1892	1902	269	<800
9519	9526	1120	1140	1612	1625
2181	2036	484	<800	588	<800
9684	9526	849	853	1023	1062
11296	11217	2509	2650	2971	2959
234	<800	41	<800	1372	1349
6265	6174	4610	4668	4124	4076
3585	3775	6930	6545	1107	1116
601	<800	1414	1429	228	<800
2655	2672	6608	6545	5400	5484
2683	2872	4638	4668	2274	2443
1588	1568	3497	3506	2413	2443
2500	2672	607	<800	1239	1236
1244	1213	6733	6545	845	850
432	<800	2765	2790	6	<800
3494	3531	3037	3069	6999	7078

```
Query Match      17.9%; Score 145; DB 9; Length 173288;
Best Local Similarity 100.0%; Pred. No. 1.5e-27;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	Db
499	CCAGAGGAGGGAACGTGTGAAGTAGCAGCACACCGATGTTGTAACAAGATTCGCATT 558
76379	CCAGAGGAGGGAACGTGTGAAGTAGCAGCACACCGATGTTGTAACAAGATTCGCATT 76438

559 GAGGAGCGGTCAAAACAGTAAAGTGTTCCTGCTACCTGGAAAAGTGGCTGGAAACA 618

Db 76439 GAGGAGCGGTCA CAAACAGTAAAGTGTTCCTGTCTTACCTGGAAAAAGTGGCTGGAAACAACA 76498

619 AGAACC GGCTTCTTGGCGTCGATG 643

Db 76499 AGAAACCGCCTTCTTGGCTGATG 76523

RESULT 13

AC112214			
LOCUS	AC112214	173990 bp	DNA linear PRI 29-MAR-2002

LOCUS	CLONING VECTOR	SEQUENCE DETERMINED	REFERENCE
DEFINITION	Homo sapiens chromosome 3 clone RP11-649F12, complete sequence.		

ACCESSION	AC112214
VERSION	AC112214.2
	GI:19807861

SOURCE	Homo sapiens (human)
REMARKS	

ORGANISM

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE	1 (bases 1 to 173990)
AUTHORS	Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saepulhimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 173990)
AUTHORS	Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE	Direct Submission
JOURNAL	Submitted (20-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE	3 (bases 1 to 173990)
AUTHORS	Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saepulhimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
TITLE	Direct Submission
JOURNAL	Submitted (29-MAR-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT	On Mar 29, 2002 this sequence version replaced g1:1676513.

9522	9507	1107	1112	484	<800
2248	2261	4124	4049	1120	1122
883	898	1371	1398	1889	1920
4124	4106	2971	3081	9661	9610
819	813	1023	979	525	<800
3304	3334	588	<800	315	<800
1141	1109	1612	1612	4194	4146
4677	4566	269	<800	562	<800
1420	1368	2160	2253	945	956
1404	1368	10528	10505	5463	5476
20	<800	3272	3197	2869	2869
7691	7607	3107	3197	1956	2031
109	<800	829	849	3465	3478
123	<800	1131	1112	1694	1690
2440	2442	2106	2033	271	<800
2429	2553	2619	2622	10687	10712
2168	2145	1419	1398	7222	7247
517	<800	996	979	6205	6158
1101	1109	2443	2409	3064	3027
1459	1368	6429	6428	2177	2189
314	<800	7603	7648	1688	1690
4210	4106	3560	3557	46	<800

Query Match 17.9%; Score 145; DB 9; Length 173990;
 Best Local Similarity 100.0%; Pred. No. 1.5e-27;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
 AY325123 901 bp mRNA linear ROD 23-JUL-2003
 LOCUS AY325123
 DEFINITION Mus musculus TAFA4 mRNA, complete cds.
 ACCESSION AY325123
 VERSION AY325123.1 GI:32967244
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 901) Tang, Y.T., Emage, P., Funk, W., Hu, T., Atterburn, M., Park, E. and Rupp, F.
JOURNAL	TAFA: A Novel Secreted Family with Homology to CC-chemokines
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 901) Tang, Y.T., Emage, P., Funk, W., Hu, T., Atterburn, M., Park, E. and Rupp, F.
JOURNAL	Direct Submission
FEATURES	Submitted (17-JUN-2003) Biology Research, Nucleo, 675 Almanor Avenue, Sunnyvale, CA 94085, USA
source	Location/Qualifiers 1..901 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090"
CDS	206..613 /note="secreted protein similar to CC-chemokines" /codon_start=1 /product="TAFA4" /protein_id="AAP92415.1" /db_xref="GI:32967245" /translation="MRVCAKVLISRLVLYIVMCKMSASSQHLRGHAGHLIK PGCEVAVHRCNRIIEERQIVKSCFPGQVAGTRAPQSPVEALIVLRKWMCM NPLEGEDCKVLPDSSGWSGSGNKKVTKTKVR"

ORIGIN

Query Match	16.9%; Score 137.4; DB 10; Length 901;
Best Local Similarity	67.0%; Pred. No. 1.5e-25;
Matches	195; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY	507	AGGAGGTGTGAAGTATGACGACACCGATGTTGTAACAAGATTCGATGAGAGCG	566
DB	340	AGGTACCTCGAGGTGTGCGCGTGCACCGCTGCAATTAAGAACCGATAGAGAGCG	399
QY	567	GTCAACAACGTAAAGTGTCTGCTACTACCTGGAAGTGGCTGGAACAACAAGAACCG	626
DB	400	CTCTCAGAGGTCAATGTTCTGCTCCAGGCCAGCGTGGCGACAACTCGGCACA	459
QY	627	GCCTTTGGCTGATGCTCCATGATGATGGAATGATGGAAATGATGATGAGAGCTTG	686
DB	460	GCCTTTGTTGTGATGAGTGCATGTTGATGAGAAATGATGATGATGATGATGATG	519
QY	687	CCTAAGAGGAAGATGTAAGACACTCCCTGACATTTGATGATGATGATGATGATG	746
DB	520	CTTGAAGAGGAGACTGTAAAGTCTTCCAGACTCTTGGGTTGATCCTGTAGCAGTG	579
QY	747	CACAAATTAAGACGACGAAATTCACCCAGAACCTTAACAAGATTT	797
DB	580	CAATTAAGTAAACCAACCAAGTGACCCGGTAGCAGACAAAGATGTTT	630

RESULT 15

AK057890	1897 bp	mRNA	linear	PRI 12-SEP-2003
LOCUS	AK057890			
DEFINITION	Homo sapiens cDNA FLJ25161 fls, clone CBR08081.			
ACCESSION	AK057890			
VERSION	AK057890.1 GI:16553863			
KEYWORDS	oligo capping; fls (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1	Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,			
	Fukuzumi, Y., Fujimori, Y., Komiyama, M., Suzuki, Y., Hata, H.,			
	Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T.,			
	Irie, R., Otsuki, T., Sato, H., Nishikawa, T., Sugiyama, A.,			
	Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished			

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 1897)
Sugano, S. and Suzuki, Y.
Submitted (24-Oct-2001) Sunio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center, Shirokane-da, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail: flicdn@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan. cDNA full insert sequencing:
Research Association for Biotechnology (RAB), cDNA library
construction and 5'-end one pass sequencing; Institute of Medical
Science, University of Tokyo, Laboratory of Genome Structure, Human
Genome Center; 3'-end one pass sequencing: RAB; clone selection for
full insert sequencing: RAB and Helix Research Institute.

FEATURES

Source

1..1897
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CBR08081"
/tissue_type="Brain"
/clone_id="CBR"
/note="Cloning vector: pME18SFL3"
110..532
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71606.1"
/db_xref="GI:16553864"
/translation="MRSPPRYCAKSVLISHWFLAVYIMVCKLMSASQHLRHAG
HHOIKGTCEVVAHRCCKNRIBERSQTVCSPPGQVAGTPRAQPSCVASIVIK
MKCHMPLIEGSDCVLPDYSGMSSGKXKTKVTR"

CDS

ORIGIN

Query Match 16.8%; Score 136.8; DB 9; Length 1897;
Best Local Similarity 70.4%; Pred. No. 2.2e-25;
Matches 183; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

QY 507 AGGAGCGTGTAGAGATGACGACACCGATGTTGTAACAAGATCGCATTGAGAGCG 566
    |||||
DB 259 AGGAGCGTGTAGAGATGACGACACCGATGTTGTAACAAGATCGCATTGAGAGCG 318
    |||||
QY 567 GTCACAAACAGTAAGTGTCTCTGCTACCTGAAAAGTGGCTGSAACAACAAGAACG 626
    |||||
DB 319 GTCACAAACAGTCAAGTGTCTCTGCTCCGGGACAGGTGGCGGACAACTCGGGCTCA 378
    |||||
QY 627 GCCTTCTTGCTGATGCTCATAGTGTGGAATGATGATGAGATGAGACCTTG 686
    |||||
DB 379 ACCTTCTTGTGTAGAGCTTCCATGTGTATTCAGAAATGATGATGATGATGATGAT 438
    |||||
QY 687 CCTAAGAGAGAGAGATGTAGACACTCCCTGACAAATTCGTGATGATGTGGCAACAG 746
    |||||
DB 439 TTGGAAGAGAGAGATGTAAAGTGTGCGAGATTACTAGGTTGTCTGTAGCAGTGG 498
    |||||
QY 747 CAACAAATTAAGACACGA 766
    |||||
DB 499 CAATAAAGTCAAAACTACGA 518
    |||||

```

Search completed: April 9, 2004, 14:55:30
Job time : 3299 secs

Tue Apr 13 09:46:06 2004

us-09-763-335-4.rml

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 12:56:28 ; Search time 94 Seconds
(without alignments)
4793.831 Million cell updates/sec

Title: US-09-763-335-4

Perfect score: 812
Sequence: 1 caccgagctgggagatgctcc.....catttgcgtgtaagaa 812

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	131.2	16.2	716	US-09-620-312D-225
2	88.8	10.9	333	US-08-905-223-218
3	46.4	5.7	7218	US-08-232-463-14
4	44.4	5.5	832	US-09-621-976-2813
5	39	4.8	832	US-09-621-976-2813
6	38.6	4.8	2060	US-09-370-807-1
7	38.6	4.8	2060	US-09-921-259-1
8	37	4.6	2914	US-08-454-097-11
9	37	4.6	2914	US-08-185-359-11
10	37	4.6	6866	US-10-204-708-19
11	36.2	4.5	2143	US-09-673-395A-82
12	36.2	4.5	3174	US-09-489-847-86
13	36.2	4.5	580073	US-08-545-528D-1
14	35.8	4.4	472	US-09-439-313-351
15	35.8	4.4	472	US-09-352-616A-351
16	35.8	4.4	472	US-09-635-215-351
17	35.8	4.4	472	US-09-685-166A-351
18	35.4	4.4	4147	US-09-688-078-6
19	35	4.3	1000	US-09-641-638-357
20	35	4.3	5177	US-09-568-407-2
21	34.6	4.3	406	US-09-702-705-1150
22	34.6	4.3	406	US-09-736-457-1150
23	34.6	4.3	406	US-09-614-134B-1150
24	34.6	4.3	406	US-09-671-325-1150
25	34.6	4.3	5445	US-09-689-065B-2
26	34.6	4.3	640681	US-09-790-988-1
27	34.4	4.2	696	US-09-134-001C-1287

28	34.2	4.2	571	US-08-858-207A-262	Sequence 262, App
29	34.2	4.2	3867	US-09-366-715-5	Sequence 5, Appl1
30	34.2	4.2	5365	US-08-961-527-77	Sequence 77, Appl
31	33.8	4.2	1001	US-09-641-638-356	Sequence 356, App
32	33.8	4.2	3824	US-09-326-529-3	Sequence 3, Appl1
33	33.8	4.2	13086	US-08-956-171B-16	Sequence 16, Appl
34	33.8	4.2	162450	US-09-345-882-1	Sequence 1, Appl1
35	33.4	4.1	301	US-09-439-313-263	Sequence 263, App
36	33.4	4.1	301	US-09-439-313-302	Sequence 302, App
37	33.4	4.1	301	US-09-352-616A-263	Sequence 263, App
38	33.4	4.1	301	US-09-352-616A-302	Sequence 302, App
39	33.4	4.1	301	US-09-232-149A-263	Sequence 263, App
40	33.4	4.1	301	US-09-232-149A-302	Sequence 302, App
41	33.4	4.1	301	US-09-159-812-263	Sequence 263, App
42	33.4	4.1	301	US-09-159-812-302	Sequence 302, App
43	33.4	4.1	301	US-09-636-215-263	Sequence 263, App
44	33.4	4.1	301	US-09-636-215-302	Sequence 302, App
45	33.4	4.1	301	US-09-685-166A-263	Sequence 263, App

ALIGNMENTS

```
RESULT 1
US-09-620-312D-225
; Sequence 225, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Aseundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yongdong
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Zhutui
; APPLICANT: Wang, Zhutui
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radote T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/489,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_genes Version 1.0
; SEQ ID NO 225
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(596)
US-09-620-312D-225

Query Match 16.2%; Score 131.2; DB 4; Length 716;
Best Local Similarity 63.1%; Pred. No. 9,8e-32;
Matches 202; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

CY 479 AGCAGCATCCTGTCACAGACGAGAGGAGCGTGTAAAGTATGACAGACACCGAT 538
DB 295 ACATCATTAAGCTCACCATGTATTAACGGGAACCTTGAGGTGGTGCACATCCACAGAT 354
```

QY 539 GTTGTAAACAGATGCGATTGAGAGCGGCTCAAAACAGTAAGTTCCTGTACTG 598
| | | | |
Db 355 GGTGTAATGAACAGATGAGAACGGTCAAAACAGTCAAGTCTCCCTG 414
| | | | |
QY 599 GAAAGTGGCTGAACAACAAGAACCGGCTTGGTGCATGCCCTCAGTATGATTG 658
| | | | |
Db 415 GGCAGGTGGCAGGACACGACGAGCTGCTCCATCATGTGTGATGCTTCAATAGTGAC 474
| | | | |
QY 659 GGAATGTGTGTGATGAGCGCTTGCTAGAAAGAGATGTAAGACACTCCCTG 718
| | | | |
Db 475 AGAAATGGTGGTCCATATGACGACATGTCTAGAGGAGAAAGTAAGTCTCCG 534
| | | | |
QY 719 ACAAATTCGATGATGTGCGCAACAGGACAAATTAAGCCGAGAAATCAACCA 778
| | | | |
Db 535 ATCGGAAAGATGGAGCTGTCTCTGGGAATAAAGTCAAAACACTAGGGAACCAATT 594
| | | | |
QY 779 GAACCTAACAGAGCAATTG 798
| | | | |
Db 595 AACCCAGGAGAAATCAAGT 614
| | | | |
RESULT 2
US-08-905-223-218
Sequence 218, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: 32..331
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..300
OTHER INFORMATION: id R13004
FEATURE:
| | | | |

NAME/KEY: other
LOCATION: 114..274
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 54..214
OTHER INFORMATION: id T80337
FEATURE:
NAME/KEY: other
LOCATION: 272..331
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 213..272
OTHER INFORMATION: id T80337
FEATURE:
NAME/KEY: other
LOCATION: 66..106
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 6..46
OTHER INFORMATION: id T80337
FEATURE:
NAME/KEY: other
LOCATION: 101..278
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 70..247
OTHER INFORMATION: id T08840
FEATURE:
NAME/KEY: other
LOCATION: 33..113
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..81
OTHER INFORMATION: id T08840
FEATURE:
NAME/KEY: other
LOCATION: 101..249
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 95
OTHER INFORMATION: region 72..220
OTHER INFORMATION: id HSC0CF041
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 31..112
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 1..82
OTHER INFORMATION: id HSC0CF041
OTHER INFORMATION: est
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 247..321
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 5.8
OTHER INFORMATION: seg LWSACMLCHG/SL
US-08-905-223-218
Query Match 10.9%; Score 88.8; DB 3; Length 333;
Best Local Similarity 97.8%; Pred. No. 2e-18;
Matches 90; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 378 TCAGAAATGGCAATGCTCTGCGATGCTCGGCTCCTGTTTGTGATAGGCTTG 437
| | | | |
Db 240 TTAATGAATGGCAATGCTCTGCGATGCTCGGCTCCTGTTTGTGATAGGCTTG 299
| | | | |
QY 438 TGCATGCTACTCTGCGATGATCCCTGAGC 469
| | | | |

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; INFORMATION

APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXPV VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

UNRECORDED ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA

```

? ZIP: 22313-0299
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
?
? COMPUTER: IBM PC compatible
?
? OPERATING SYSTEM: PC-DOS/MS-DOS
?

```

```

1 SOFTWARE: Patent In Release #1.0, Version #1
2
3 CURRENT APPLICATION DATA:
4
5 APPLICATION NUMBER: US/08/232,463
6
7 FILING DATE:
8
9 CLASSIFICATION: 435
10
11 PRIOR APPLICATION DATA:
12

```

APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30422/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)835-9300
FAX: (703)835-9300

TELEFAX: (703) 683-4109
TELEX: 899124
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTz9pt-F15
US-08-232-463-14

```

Query Match          5.7%; Score 46.4; DB 1
Best Local Similarity 2.4%; Pred. No. 0.0003;
Matches      8; Conservative 195; Mismatches 13

```

QY 715 AGAGATACCTGCACACGACGAAGAGGGCGTGG
Db 1449 AGAAGAATTGGTACRRRRRRRRRRRRRRRRR
QY 539 GTTGTAACAAGATCCCATTTAGAGAGCGGTCAAAAC

Ddb	1389 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Qy	599 GAAATGCTGGACACAGAACCGGCTTGTG ::: 1329 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR

QY	659	GGAAATGTCGTGTGAGATGAGCCTTGCCCTGAAAG
	1269	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db	719	AAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

RESULT 4
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:

1 APPLICANT: Dumas Milne Edwards, J.B.
 2 APPLICANT: Jobert, S.
 3 APPLICANT: Giordano, J.Y.
 4 TITLE OF INVENTION: ESTs and Encoded Human Proteins
 5 FILE REFERENCE: GENSET_054PR2
 6 CURRENT ABSTRACT NUMBER: nc/00/671 076

```

? CONSECUTIVE FILLS: 1
? CONSECUTIVE FILLS: 1
? CURRENT FILING DATE: 2000-07-21
? NUMBER OF SEQ ID NOS: 19335
? SOFTWARE: Patent.pm
? SEQ ID NO 2813
? LENGTH: 832

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 235..399

```

Query Match	5.5%	Score 44.4	DB 4
Best Local Similarity	11.4%	Pred. No. 0.00044	
Matches 27; Conservative	119; Mismatches	90	

Oy	57	CACATTATCCCATCACTTCAAAGGTCGTCAAGC : :: : : : :
Dd	237	CATWCWYWKVKRMWSKTCSGRGIMSYSTRSY
Ox	117	TTPAAGTAAATGACAGA GTTTCAACCCTCAACC

Db	177	WTMAGKMKWRATTPRRAMMMMAATWMMYMMWANC
Qy	177	AGCTGAATGAACCGATGCTTTTTTTTACTGTGA

Db 117 RKKSYRRRCMAAYAMKTKBSLYYCNFKMKIKRRCMMNNNN
Qy 237 ATTTTTTTTTTAAATCCTGATAGAGAGATTGTTGGAA
57 WAMAMRWMTTMMNNNNYYVYAMKRRMMWRKMSWSNNNN

RESULT 5
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063

```

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Protein
; FIRST INVENTOR: CHANCE

```

```

1 FILE REFERENCE: GENSET.054PR2
2
3 CURRENT APPLICATION NUMBER: US/09/621,976
4
5 CURRENT FILING DATE: 2000-07-21
6
7 NUMBER OF SEQ ID NOS: 19335
8
9 SOFTWARE: Patent.pm
10
11 SEQ ID NO 2813

```

```

?      LENGTH: 832
?
?      TYPE: DNA
?
?      ORGANISM: Homo sapiens
?
?      FEATURE:
?
?      NAME/KEY: CDS

```


REFERENCE/DOCKET NUMBER: 27866/31853
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 265..1275
US-08-454-097-11

Query Match 4.6%; Score 37; DB 1; Length 2914;
Best Local Similarity 50.9%; Pred. No. 0.22;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 181 TGAATGAACCCGATGCTTTTCTTTTCTGGAATAGATCGGAGAGATACATTT 240
DB 2193 TGAGTCAGCAGGCAATTTTGTGTTTAAAAAATCTCATTCTTACGAAACAGTT 2134
QY 241 TTTTCTTAATCTGATPAAAGAAAGATTGTGGAGCTCTTTGAAAAAATTTTAAATTG 300
DB 2133 TTTAGTTTAAATGAAGCTGTAAACAAAAGCTCCATTCAAAATPAAACAAATCC 2074
QY 301 TGGCAGATGATGATTTTAAAGTGTAGATCTTCCATGATGACACTATGA 353
DB 2073 CAGATCATATGATGTTTACAGTGTATCATTTATGAGCAACATCATCA 2021

RESULT 9
US-08-185-359-11/C
Sequence 11, Application US/08185359
Patent No. 6060296
GENERAL INFORMATION:
APPLICANT: Hoechst, Merl F.
TITLE OF INVENTION: Protein Kinases
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,359
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,001
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,783
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 6060296and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31853
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 25-3856
INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 265..1275
US-08-185-359-11

Query Match 4.6%; Score 37; DB 3; Length 2914;
Best Local Similarity 50.9%; Pred. No. 0.22;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 181 TGAATGAACCCGATGCTTTTCTTTTCTGGAATAGATCGGAGAGATACATTT 240
DB 2193 TGAGTCAGCAGGCAATTTTGTGTTTAAAAAATCTCATTCTTACGAAACAGTT 2134
QY 241 TTTTCTTAATCTGATPAAAGAAAGATTGTGGAGCTCTTTGAAAAAATTTTAAATTG 300
DB 2133 TTTAGTTTAAATGAAGCTGTAAACAAAAGCTCCATTCAAAATPAAACAAATCC 2074
QY 301 TGGCAGATGATGATTTTAAAGTGTAGATCTTCCATGATGACACTATGA 353
DB 2073 CAGATCATATGATGTTTACAGTGTATCATTTATGAGCAACATCATCA 2021

RESULT 10
US-10-204-708-19
Sequence 19, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PREPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 19
LENGTH: 6866
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-19

Query Match 4.6%; Score 37; DB 4; Length 6866;
Best Local Similarity 51.5%; Pred. No. 0.37;
Matches 85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 186 GAACCCGATGCTTTTCTTTTCTGGAATAGATCGGAGAGATACATTTT 245
DB 3859 GAATAGATTAATGTTTCTTTTAAATAGTAATAGTAATATTAATGAAAT 3918
QY 246 TTAATCTGATPAAAGAAAGATTGTGGAGCTCTTTGAAAAAATTTTAAATTG 305
DB 3919 TTAATGATTAATPAAAGAAAGATTGTGAAAGAAATATTAATGAGGTTGT 3978
QY 306 CAGATGATTTTAAAGTGTAGATCTTCCATGATGACACTAT 350

Db 3979 TAAATGTTAATGATTAAGTTGGGAGTTTAAATTAGGATTTAAT 4023

RESULT 11

US-09-673-395A-82/c
; Sequence 82, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMAN, BERNI
; APPLICANT: SCHWITT, ARMIN
; APPLICANT: PILARSKI, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 82
; LENGTH: 2143
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-673-395A-82

Query Match 4.5%; Score 36.2; DB 4; Length 2143;
Best Local Similarity 52.3%; Pred. No. 0.34;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 224 GGAGAGAGTACATTTTCTTATCCTGATTAAGATGTTGGAGCTCTTGA 283
DB 736 GGAATACATGGTGTGTTTATCATAGTGCACATTTAGTTGGTAATGACTACA 677
QY 284 AAAAAATTTTAATGTGGCAGATGATTTAAAAAGTTAGTCTTCCATGAA 343
DB 676 AAACGATTTTAAATCAAGTTAATGTAATTTGAAATTAACCTTAATCTCA 617
QY 344 CACTAATAGACTCTGCTGCTGGCTGATTT 376
DB 616 CATAACATGGCATTAAGTTTGACTTGAGTT 584

RESULT 12

US-09-489-847-86
; Sequence 86, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: ROSEN, et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 86
; LENGTH: 3174
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-86

Query Match 4.5%; Score 36.2; DB 4; Length 3174;
Best Local Similarity 52.3%; Pred. No. 0.43;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 224 GGAGAGAGTACATTTTCTTATCCTGATTAAGATGTTGGAGCTCTTGA 283
DB 2285 GGAATACATGGTGTGTTTATCATAGTGCACATTTAGCTTGTGTAATGACTACA 2344
QY 284 AAAAAATTTTAATGTGGCAGATGATTTAAAAAGTTAGTCTTCCATGAA 343
DB 2345 AAACGATTTTAAATCAAGTTAATGTAATTTGAAATTAACCTTAATCTCA 2404
QY 344 CACTAATAGACTCTGCTGCTGGCTGATTT 376
DB 2405 CATAACATGGCATTAAGTTTGACTTGAGTT 2437

RESULT 13

US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: FRASER et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragments
; Patent No. 6537773
; FILE REFERENCE: P8193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/489,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 4.5%; Score 36.2; DB 4; Length 580073;
Best Local Similarity 56.2%; Pred. No. 8.8;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 199 TTTTCTTACTGTGGAATAGATCGAAGAGATMACTTTTCTTATCCTGATTA 258
DB 502669 TTAATTTTCTGTATCAACAAAAGAGCAATATTGCTGCTTAAATGTTATTA 502610
QY 259 AGAAGATTGTGGAGAGCTCTTGAATAATTTTAATGTGGCAGATGATTTTA 318
DB 502609 AAAAAACAGAGCTTAAGATTTTACAGAAAATTTTGTCTGTATGATGATATTA 502550
QY 319 A 319
DB 502549 A 502549

RESULT 14

US-09-439-313-351/c
; Sequence 351, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: XU, Jiangchun
; APPLICANT: DILLON, Davin C.
; APPLICANT: MITCHELL, Jennifer L.
; APPLICANT: HARLOCKER, Susan Louise
; APPLICANT: JIANG YUGU
; APPLICANT: REED, Steven G.
; APPLICANT: KALOS, Michael
; APPLICANT: FANGER, Gary
; APPLICANT: RETTER, Mark

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 15:02:33 ; Search time 355 Seconds

(without alignments)
8580,808 Million cell updates/sec

Title: US-09-763-335-4

Perfect score: 812

Sequence: 1 caccggagctgggagatgctcc.....catttgrgtaagaagaa 812

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2475585 seqs, 187570760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCT05_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.2	16.2	716	US-10-037-270-225	Sequence 225, App
2	131.2	16.2	716	US-10-117-722-225	Sequence 225, App
3	128.4	15.8	2390	US-10-104-047-892	Sequence 892, App
4	43.4	5.3	3673778	US-10-312-841-1	Sequence 1, App
5	42.6	5.2	16724	US-10-311-455-1064	Sequence 1064, App
6	42.6	5.2	16724	US-10-240-485-90	Sequence 90, App
7	42	5.2	498	US-09-991-936-1045	Sequence 1045, App
8	41.8	5.1	10886	US-10-311-455-2108	Sequence 2108, App
9	41.4	5.1	40862	US-10-311-455-2046	Sequence 2046, App
10	41	5.0	17869	US-10-311-455-78	Sequence 78, App
11	41	5.0	17869	US-10-257-166-3	Sequence 2, App
12	40.4	5.0	13511	US-10-311-455-253	Sequence 253, App
13	40.2	5.0	9268	US-10-311-455-1573	Sequence 1573, App
14	40.2	5.0	3673778	US-10-312-841-2	Sequence 2, App
15	39.8	4.9	13038	US-10-311-455-1248	Sequence 1248, App

16	39.6	4.9	5875	US-10-311-455-262	Sequence 262, App
17	39.6	4.9	11416	US-10-221-613-19	Sequence 19, App
18	39.6	4.9	11416	US-10-311-455-91	Sequence 91, App
19	39.4	4.9	89328	US-09-873-367C-332	Sequence 332, App
20	39.2	4.8	6904	US-10-311-455-186	Sequence 186, App
21	39.2	4.8	23934	US-09-764-877-2455	Sequence 2455, App
22	39.2	4.8	23934	US-10-242-515-2455	Sequence 2455, App
23	39	4.8	7276	US-10-311-455-875	Sequence 875, App
24	38.8	4.8	8979	US-10-221-613-138	Sequence 138, App
25	38.8	4.8	8979	US-10-311-455-758	Sequence 758, App
26	38.6	4.8	5269	US-10-311-455-2029	Sequence 2029, App
27	38.2	4.7	652	US-10-125-968-1355	Sequence 1355, App
28	38.2	4.7	16914	US-10-221-613-214	Sequence 214, App
29	38	4.7	10716	US-10-311-455-1391	Sequence 1391, App
30	37.8	4.7	564	US-10-027-632-286621	Sequence 286621, App
31	37.8	4.7	564	US-10-027-632-286628	Sequence 286628, App
32	37.8	4.7	564	US-10-027-632-286629	Sequence 286629, App
33	37.8	4.7	564	US-10-027-632-286630	Sequence 286630, App
34	37.6	4.6	14253	US-10-311-455-1467	Sequence 1467, App
35	37.4	4.6	5970	US-10-108-605-210	Sequence 210, App
36	37.4	4.6	6282	US-10-108-605-212	Sequence 212, App
37	37.4	4.6	7049	US-10-311-455-129	Sequence 129, App
38	37.4	4.6	7049	US-10-240-452-5	Sequence 5, App
39	37.4	4.6	9155	US-10-311-455-435	Sequence 435, App
40	37.4	4.6	17848	US-10-235-676-28	Sequence 28, App
41	37.4	4.6	17848	US-10-240-453-38	Sequence 38, App
42	37.4	4.6	17848	US-10-257-166-58	Sequence 58, App
43	37.2	4.6	8136	US-10-311-455-527	Sequence 527, App
44	37.2	4.6	8136	US-10-257-166-37	Sequence 37, App
45	37	4.6	532	US-09-814-353-13846	Sequence 13846, App

ALIGNMENTS

RESULT 1
US-10-037-270-225
Sequence 225, Application US/10037270
Publicatation No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundang
APPLICANT: Wang, Duntui
APPLICANT: Wang, Zhilwei
APPLICANT: Tillinghast, John
APPLICANT: Dermanac, Radote T.
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT FILING DATE: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pt_Fl_genes Version 1.0
SEQ ID NO 225
LENGTH: 716
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:


```
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (3294164)
US-10-312-841-1

Query Match
Best Local Similarity 5.3%; Score 43.4; DB 14; Length 3673778;
Matches 104; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 239 TTTTTCCTGATTAAGAGATGCTGGAGAGCTCTTGAATAAATTTTAAT 298
DB 2721915 TGTATTTTATTTATTTATGAGTAGTGTGTGAGATTTTGAAGAGATTTTATAT 2721974
QY 299 TGTGGCAGATGATGATTTTAAAGTGTAGATCTTCCATGAACTAATAGAGTACT 358
DB 2721975 TTTTGTAGATGATGATTTTATGATTTTATTTTGAAGTAAATGTGAATGGAGCT 2722034
QY 359 CTGCTCTTGGCTGATTTTTCAGAGATGCGATGCTCTGCGATGCTGGGCTCTGT 418
DB 2722035 TATTATGATTTGTTTGTGTTTGTGTTATGCTATAGAGAAAGTTGTGATTTT 2722094
QY 419 ATTGTGATTAAGTGTCTGTGCAAT 443
DB 2722095 GTATATGATTTTGTATTTTGTAGAT 2722119

RESULT 5
US-10-311-455-1064
Sequence 1064, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPERROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1064
LENGTH: 16724
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1064

Query Match
Best Local Similarity 5.2%; Score 42.6; DB 14; Length 16724;
Matches 141; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 126 AAAATGACAAGTTTCCACCCCAACCTTGCTCTTCTGCAATACAGTGTGAT 185
DB 13695 AAGTTAGATGGGTATTTGATTTAGTTGGGTTATAGAGTAGATTTATTTTAA 13754
QY 186 GAACCCGATGCTTTTCTTCTGGAATATGATGCGAAGAGATGATACATTTT 245
DB 13755 AAAAATGATTAATTTTATTTTAAATTTGATTAAGATATGATGATTAATTT 13814
QY 246 TTAATCTGATTAAGAAGATGTTGGGAAGCTCTTGAATAAATTTTAAATTTG 305
DB 13815 TGGTATGTTTAAAGATGTTGATTTGTTTGAAGAATAATTTATGATTTGA 13874
QY 306 CAGATGATTTTAAAGATGTTGATCTTCCATGAACTAATAGAGTACTGCTCT 365
DB 13995 AAAAA 13999
```

```
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (3294164)
US-10-312-841-1

Query Match
Best Local Similarity 5.2%; Score 42.6; DB 14; Length 16724;
Matches 141; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 126 AAAATGACAAGTTTCCACCCCAACCTTGCTCTTCTGCAATACAGTGTGAT 185
DB 13695 AAGTTAGATGGGTATTTGATTTAGTTGGGTTATAGAGTAGATTTATTTTAA 13754
QY 186 GAACCCGATGCTTTTCTTCTGGAATATGATGCGAAGAGATGATACATTTT 245
DB 13755 AAAAATGATTAATTTTATTTTAAATTTGATTAAGATATGATGATTAATTT 13814
QY 246 TTAATCTGATTAAGAAGATGTTGGGAAGCTCTTGAATAAATTTTAAATTTG 305
DB 13815 TGGTATGTTTAAAGATGTTGATTTGTTTGAAGAATAATTTATGATTTGA 13874
QY 306 CAGATGATTTTAAAGATGTTGATCTTCCATGAACTAATAGAGTACTGCTCT 365
DB 13995 AAAAA 13999

Query Match
Best Local Similarity 46.2%; Pred. No. 0.85;
Matches 141; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 126 AAAATGACAAGTTTCCACCCCAACCTTGCTCTTCTGCAATACAGTGTGAT 185
DB 13695 AAGTTAGATGGGTATTTGATTTAGTTGGGTTATAGAGTAGATTTATTTTAA 13754
QY 186 GAACCCGATGCTTTTCTTCTGGAATATGATGCGAAGAGATGATACATTTT 245
DB 13755 AAAAATGATTAATTTTATTTTAAATTTGATTAAGATATGATGATTAATTT 13814
QY 246 TTAATCTGATTAAGAAGATGTTGGGAAGCTCTTGAATAAATTTTAAATTTG 305
DB 13815 TGGTATGTTTAAAGATGTTGATTTGTTTGAAGAATAATTTATGATTTGA 13874
QY 306 CAGATGATTTTAAAGATGTTGATCTTCCATGAACTAATAGAGTACTGCTCT 365
DB 13995 AAAAA 13999
```

RESULT 7

US-09-991-936-1045/c
; Sequence 1045, Application US/09991936
; Publication No. US20030072827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1045
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1045

Query Match 5.2%; Score 42; DB 10; Length 498;
Best Local Similarity 47.9%; Pred. No. 0.16; Indels 0; Gaps 0;

Matches 105; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 164 TTTCGACATACAGTCTGATGACACCGATGCTTTTCTGGAATAGATC 223
DB 486 TTTTAAANTGANNCTGGAATGATGATGAATGAAATATTAATCTTAAATTA 427
QY 224 GGAAGAGAGTACATTTTCTGATGATGATGATGATGATGATGATGATGAT 283
DB 426 GTTATGATATTAATTTTATGTTAAAGGCTAAATTTTAAAGCGAAGACCT 367
QY 284 AAAAAATTTTAAATTTGCGACAGATGATTTTAAAGTGTAGATCTTCCAAATGA 343
DB 366 TATTAAGTTTA 307
QY 344 CACTAATAGAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382
DB 306 TTATATACATTTTATTTGGGCTGATGAGANAATTTTATATAA 268

RESULT 8

US-10-311-455-2108
; Sequence 2108, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2108
; LENGTH: 10886
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-2108

Query Match 5.1%; Score 41.8; DB 14; Length 10886;
Best Local Similarity 50.2%; Pred. No. 1.1; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 239 TTTTCTTTTATCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 298
DB 10668 TCTTTTCTTTTATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 10727
QY 299 TGTGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 10728 TTTTGTAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10787
QY 359 CTGCTCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 418
DB 10788 TATTTATGATTTGGTCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 10847
QY 419 ATTTGTGATAGTCTGTCAT 443
DB 10848 GTATATGATTTTGTATTTTGAT 10872

RESULT 9

US-10-311-455-2046
; Sequence 2046, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2046
; LENGTH: 40862
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2046

Query Match 5.1%; Score 41.4; DB 14; Length 40862;
Best Local Similarity 57.3%; Pred. No. 3.2; Indels 0; Gaps 0;
Matches 75; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 193 ATGCTCTTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
DB 1414 ATGCTCTTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1473
QY 253 TGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 312
DB 1474 TTTTAAGTAAATATGAGAAATTTTAAATGATTTTAAATGATTTTAAATGATGAT 1533
QY 313 ATTTTAAAG 323
DB 1534 AATTGAATAG 1544

RESULT 10

US-10-311-455-78
; Sequence 78, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:

Tue Apr 13 09:46:06 2004

us-09-763-335-4.rnpb

Page 7

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 12:42:37 ; Search time 2543 Seconds

(without alignments)
9535.230 Million cell updates/sec

Title: US-09-763-335-4
Perfect score: 812
Sequence: 1 cactgagtgaggagtgatcc.....cattgtgtagtaagaa 812

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc1:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc1:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	812	100.0	1934	11	BC025746 Homo sapi
2	800.2	99.5	929	9	AL534074 AL534074
3	767.2	94.5	940	13	BX460561 BX460561
4	667	82.1	793	12	BT756155 BT756155

Result No.	Score	Query Match	Length	ID	Description
5	610.2	75.1	1022	12	BT488689
6	551.6	67.9	3258	11	AK050001
7	511	62.9	672	14	CK001752
8	475.4	58.5	768	13	BU118807
9	442.2	54.5	506	10	BF652347
10	440.8	54.3	706	13	BU615139
11	434.8	53.5	975	9	AL533965
12	409	50.4	468	14	H23443
13	389.2	47.9	692	10	BB653506
14	366.6	45.1	2412	11	AK049880
15	361.8	44.6	889	13	BU275101
16	353.4	43.5	845	13	BU272791
17	349	43.0	728	10	AM955725
18	346.8	42.7	443	14	R67171
19	320.4	39.5	442	14	R56289
20	318.4	39.2	377	14	T81284
21	294	36.2	853	13	BU281041
22	276	34.0	357	28	B2950260
23	274.6	33.8	472	14	CB730363
24	257.6	31.7	473	10	BA365782
25	251	30.9	251	9	AA365782
26	215	26.5	420	14	T80333
27	201	24.8	594	10	AM641769
28	200	24.6	558	10	BB653101
29	199.4	24.6	510	9	AL861103
30	199.4	24.6	643	9	AL861104
31	197.8	24.4	406	13	BY283942
32	187.6	23.1	451	13	BY256413
33	181.6	22.4	484	9	AM049606
34	173.8	21.4	436	13	BY283473
35	173.8	21.4	894	13	BU133938
36	164.8	20.3	460	14	R13004
37	145	17.9	538	28	AO318831
38	144.2	17.8	440	13	BY250323
39	143.2	17.6	778	13	EX101385
40	139.4	17.2	535	29	CG561270
41	136.8	16.8	397	29	AY404310
42	136.8	16.8	397	29	AY404312
43	136.8	16.8	637	13	EX085931
44	136.8	16.8	873	12	BT599941
45	135.8	16.7	4319	11	AK078681

ALIGNMENTS

RESULT 1
BC025746 1934 bp mRNA linear HTC 04-MAR-2003
LOCUS Homo sapiens, clone IMAGE:5200142, mRNA.
DEFINITION BC025746.1 GI:22213081
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (06-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;

REMARK

COMMENT

Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgti.nih.gov
 Ahlert, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stancijevic, S., Thomas, P.J., Touchman, J.W.,
 Tsirigotis, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRK Plate: 49 Row: b Column: 3
 This clone has the following problem: retained intron.

FEATURES

Location/Qualifiers
 1..1934
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5200142"
 /issue_type="Brain, adult, 6 pooled whole brains"
 /clone_lib="NTH MGC_114"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 100.0%; Score 812; DB 11; Length 1934;
 Best Local Similarity 100.0%; Pred. No. 4.7e-222; Indels 0; Gaps 0;
 Matches 812; Conservative 0; Mismatches 0;

```

1  CACTGAGATGGGAGATGTCATCGGCACTATAAATGATTCATGAGAAATGCGACA 60
Db  9  CACTGAGATGGGAGATGTCATCGGCACTATAAATGATTCATGAGAAATGCGACA 68
Qy  61  TTATCTCCCATCACTTCAAAAGTCTGTCAGGCAAGAGTGAAGCCGAGAGATATTAA 120
Db  69  TTATCTCCCATCACTTCAAAAGTCTGTCAGGCAAGAGTGAAGCCGAGAGATATTAA 128
Qy  121 AGGTAAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGAAATACAGTC 180
Db  129 AGGTAAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGAAATACAGTC 188
Qy  181 TGAATGAACCCGATGCTTTTCTGAGAAATGATCGGAGAGAGATTAATTT 240
Db  189 TGAATGAACCCGATGCTTTTCTGAGAAATGATCGGAGAGAGATTAATTT 248
Qy  241 TTTTCTTATCTGATTAAGATTTGTTGGAGCTCTTTGAAAAAATTTAAATG 300
Db  249 TTTTCTTATCTGATTAAGATTTGTTGGAGCTCTTTGAAAAAATTTAAATG 308
Qy  301 TGGCAAGATGATTTTAAAGTGTGATTTTTCATGAAACCTTAATGAGTACT 360
Db  309 TGGCAAGATGATTTTAAAGTGTGATTTTTCATGAAACCTTAATGAGTACT 368
Qy  361 GGTCTTGGTGAATTTTTCAGAAATGGCAATGCTCGGAGATCTCGGCTCTGAT 420
Db  369 GGTCTTGGTGAATTTTTCAGAAATGGCAATGCTCGGAGATCTCGGCTCTGAT 428
Qy  421 TTGTGATTAAGTCTGTGCAATGCTACTCTGCAATGATCCCTTCAGACACTTTCCAG 480
Db  429 TTGTGATTAAGTCTGTGCAATGCTACTCTGCAATGATCCCTTCAGACACTTTCCAG 488
Qy  481 CAGCATCACTGACAGACAGAGAGGAGAGTGAATGATAGACACACCGATGT 540
Db  489 CAGCATCACTGACAGACAGAGAGGAGAGTGAATGATAGACACACCGATGT 548
Qy  541 TGTAAACAAATGCTTGAAGAGCGGTCAAAACAGTAAAGTTTCTGCTACCTGA 600
Db  549 TGTAAACAAATGCTTGAAGAGCGGTCAAAACAGTAAAGTTTCTGCTACCTGA 608
Qy  601 AAAGTGGTGAACAAAGAAACCGGCTTCTTGGTGAATGCTCTCAATGATGATGG 660
  
```

Db 609 AAAGTGGTGAACAAAGAAACCGGCTTCTTGGTGAATGCTCTCAATGATGATGG 668
 Qy 661 AAAGTGGTGGATGAGATGAGACCTTGGCTGAAAGAGAGATGTAGACACTCCCTGAC 720
 Db 669 AAAGTGGTGGATGAGATGAGACCTTGGCTGAAAGAGAGATGTAGACACTCCCTGAC 728
 Qy 721 AATTGATGATGATGTCGCAACAGCAACAAATTAAGACCAAGATTCACCCAGA 780
 Db 729 AATTGATGATGATGTCGCAACAGCAACAAATTAAGACCAAGATTCACCCAGA 788
 Qy 781 ACCTAACAGAACATTTGTGTGATTAAGAA 812
 Db 789 ACCTAACAGAACATTTGTGTGATTAAGAA 820

RESULT 2
 AL534074 929 bp mRNA linear EST 12-MAY-2003
 LOCUS
 DEFINITION
 CS0DF003YGI8 5-PRIME mRNA sequence.
 ACCESSION
 AL534074 GI:30539521
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 On Feb 13, 2001 this sequence version replaced gi:12797567.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 2511.f
 Contact : Feng Liang Email : liang@life-tech.com URL :
<http://fulllength.invitrogen.com/Invitrogen/Compartion> 1600
 Faraday Avenue Genoscope sequence ID : CS0DF003BD09QPl.

FEATURES

Location/Qualifiers
 1..929
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF003YGI8"
 /issue_type="FETAL BRAIN"
 /dev_stage="Fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo (d) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."

ORIGIN

Query Match 98.5%; Score 800.2; DB 9; Length 929;
 Best Local Similarity 99.5%; Pred. No. 1e-218; Indels 1; Gaps 1;
 Matches 810; Conservative 2; Mismatches 0;

```

1  CACTGAGATGGGAGATGTCATCGGCACTATAAATGATTCATGAGAAATGCGACA 60
Db  66  CACTGAGATGGGAGATGTCATCGGCACTATAAATGATTCATGAGAAATGCGACA 125
Qy  61  TTATCTCCCATCACTTCAAAAGTCTGTCAGGCAAGAGTGAAGCCGAGAGATATTAA 120
Db  126 TTATCTCCCATCACTTCAAAAGTCTGTCAGGCAAGAGTGAAGCCGAGAGATATTAA 185
Qy  121 AGGTAAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 180
Db  186 AGGTAAATGACAAAGTTCACCCCTCAACCTTGCTCTTTTCTGACATACAGTC 245
  
```

QY 181 TGAATGAACCCGATGCTTTTCTTTTCTGAGAAATAGATGCGAAGAGATTAACATT 240
 Db 246 TGAATGAACCCGATGCTTTTCTTTTCTGAGAAATAGATGCGAAGAGATTAACATT 305
 QY 241 TTTTCTTTATCTCCTGATTAAGAAAGATGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG 300
 Db 306 TTTTCTTTATCTCCTGATTAAGAAAGATGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG 365
 QY 301 TGGACAGATGATTTTAAAAAGTTAGATCTTTCCAAATGACATTAATAGATCTCT 360
 Db 366 TGGACAGATGATTTTAAAAAGTTAGATCTTTCCAAATGACATTAATAGATCTCT 425
 QY 361 GCTCTTGCTGGATTTTTCAGAGAAATGCGAATGCTCTGCGATGCTGCTGCTGCTGCTAT 420
 Db 426 GCTCTTGCTGGATTTTTCAGAGAAATGCGAATGCTCTGCGATGCTGCTGCTGCTGCTAT 485
 QY 421 TTTGGATTAAGTCTGTTGCAATGCTACTCTGCAATGATCCCTTCAGCACTTTCCAG 480
 Db 486 TTTGGATTAAGTCTGTTGCAATGCTACTCTGCAATGATCCCTTCAGCACTTTCCAG 545
 QY 481 CAGCATCACTGCAAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 Db 546 CAGCATCACTGCAAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
 QY 541 TGTAAACAAGATTCGATTAAG 599
 Db 606 TGTAAACAAGATTCGATTAAG 665
 QY 600 AAAAGTGGCTGGAACAACAAGAAACCGGCTTTCTTGCTGATGCTGCTGCTGCTGCTGCTG 659
 Db 666 AAAAGTGGCTGGAACAACAAGAAACCGGCTTTCTTGCTGATGCTGCTGCTGCTGCTGCTG 725
 QY 660 GAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
 Db 726 GAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 785
 QY 720 CAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 779
 Db 786 CAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 845
 QY 780 AACCTAACAAGACATTTTGTGTAGTAAGAA 812
 Db 846 AACCTAACAAGACATTTTGTGTAGTAAGAA 878

RESULT 3
 LOCUS BX460561 940 bp mRNA linear EST 22-WAY-2003
 DEFINITION BX460561 Homo sapiens PITAL BRAIN Homo sapiens cDNA clone
 ACCESSION BX460561
 VERSION BX460561.1 GI:31035211
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 940)
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seque@genoscope.cns.fr Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 2511.f.
 Contact: Feng Liang Email: fliang@life.techn.com URL:
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID: CS0DF015DF120P1.
 Location/Qualifiers
 1..940
 /organism="Homo sapiens"

FEATURES
 source
 1..940
 /organism="Homo sapiens"

ORIGIN

Query Match 94.5%; Score 767.2; DB 13; Length 940;
 Best Local Similarity 97.2%; Pred. No. 3.1e-209;
 Matches 791; Conservative 10; Mismatches 9; Indels 4; Gaps 2;

QY 3 CTGAGATGGAGATGCTCCATTCGGCAACTATATAGATTCATCAGGAACTGCAATT 62
 Db 50 CCGAATTCCTCCGGATTCATTCGCAACTATATAGATTCATCAGGAACTGCAATT 109
 QY 63 ATCTCCCATCACTTCAAGAGTCTGTCAGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAG 122
 Db 110 ATCTCCCATCACTTCAAGAGTCTGTCAGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAG 169
 QY 123 GTGAAATGACAAGATTTTCAACCCCTCAACCTGCTCTTTTGTGACATACAGTCTG 182
 Db 170 GTGAAATGACAAGATTTTCAACCCCTCAACCTGCTCTTTTGTGACATACAGTCTG 229
 QY 183 AATGAACCCGATGCTTTTCTTTTCTGATGGAATAGATCGGAAAGAGTAAACA---TT 239
 Db 230 AATGAACCCGATGCTTTTCTTTTCTGATGGAATAGATCGGAAAGAGTAAACAATT 289
 QY 240 TTTTCTTTATCCTGATTAAGA-AGATGTTGGGAAGCTTTGAAAAAAATTTTAAAT 298
 Db 290 TTTTCTTTATCCTGATTAAGAAGATGTTGGGAAGCTTTGAAAAAAATTTTAAAT 349
 QY 299 TGTGACAGATGATTTTAAAAAGTGTAGATCTTCCATGAACACTTAATAGATCT 358
 Db 350 TGTGACAGATGATTTTAAAAAGTGTAGATCTTCCATGAACACTTAATAGATCT 409
 QY 359 CTGCTCTTGCTGATTTTTCAGAGAGTGAAGTGAAGTCTCTGATGCTGCTGCTGCTGCT 418
 Db 410 CTGCTCTTGCTGATTTTTCAGAGAGTGAAGTGAAGTCTCTGATGCTGCTGCTGCTGCT 469
 QY 419 ATTGTGATTAAGTCTGTTGCAATCTTCTGCAATGATGATGATGATGATGATGATGATGAT 478
 Db 470 ATTGTGATTAAGTCTGTTGCAATCTTCTGCAATGATGATGATGATGATGATGATGATGAT 529
 QY 479 AGCAGATCACTGCAAGACAG 538
 Db 530 AGCAGATCACTGCAAGACAG 589
 QY 539 GTTGTAAACAAGATTCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
 Db 590 GTTGTAAACAAGATTCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 649
 QY 599 GAAAAAGTGTGGAACAACAAGAAACCGGCTTTCTTGCTGATGATGATGATGATGATGATG 658
 Db 650 GAAAAAGTGTGGAACAACAAGAAACCGGCTTTCTTGCTGATGATGATGATGATGATGATG 709
 QY 659 GGAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 718
 Db 710 GGAATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 769
 QY 719 ACAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 778
 Db 770 ACAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 829
 QY 779 GAACCTAACAAGACATTTTGTGTAGTAAGAA 812
 Db 830 GAACCTAACAAGACATTTTGTGTAGTAAGAA 863

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /locus_tag="CS0DF015Y124"
 /clone_lib="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
 was primed with a NotI-oligo (dt) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."

RESULT 4
LOCUS B1756155
DEFINITION 793 bp mRNA linear EST 25-SBP-2001
60303005F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200142 5',
mRNA sequence.
ACCESSION B1756155
VERSION B1756155.1 GI:15747733
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 793)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rc@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11501 row: f column: 15
High quality sequence stop: 793.
Location/Qualifiers
1..793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5200142"
/lab_host="DH10B"
/clone_id="NIH_MGC_114"
/note="Organ: brain; Vector: PCMV-SPORE; Site: 1: Noci;
Site 2: EcorV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (pcov site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."

ORIGIN

Query Match 82.1%; Score 667; DB 12; Length 793;
Best Local Similarity 98.2%; Pred. No. 1.9e-180;
Matches 769; Conservative 0; Mismatches 5; Indels 9; Gaps 9;

QY 1 CACTGGAGTGGGAGTGTCCATCGCAACTATTAATGATTTCTCATCGGA-AACTGCAC 59
DB 9 CACTGGAGTGGGAGTGTCCATCGCAACTATTAATGATTTCTCATCGGAAGACTGCAC 68
QY 60 ATTATCTCCCTCACTTCAAGCTCTGTCAAGAGGAGGAGCGCCGAGAGATTTA 119
DB 69 ATTATCTCCCTCACTTCAAGCTCTGTCAAGAGGAGGAGCGCCGAGAGATTTA 128
QY 120 AAGTGAATAAGCAAGGTTTCCACCCCTCAACCTTGGCTCTTTTCTGCAATACAGT 179
DB 129 AAGTGAATAAGCAAGGTTTCCACCCCTCAACCTTGGCTCTTTTCTGCAATACAGT 188
QY 180 CTGAATGAACCCGATGTCTTTTCTGCAATAGATGCGAAGAGTAACATT 239
DB 189 CTGAATGAACCCGATGTCTTTTCTGCAATAGATGCGAAGAGTAACATT 248
QY 240 TTTTCTTATCTCTGTAAGAAGATTTGGGAAGCTTTTGAAG-AAAAATTTAAT 298
DB 249 TTTTCTTATCTCTGTAAGAAGATTTGGGAAGCTTTTGAAG-AAAAATTTAAT 308
QY 299 TGGGACAGATGATTTAAAGTGTAGATCTTCAATGAACACTAATAGAGTACT 358

DB 309 GGTGGACAGATGATTTTAAAGGTGATCTTTCATATGACACTAATAGAGTACT 368
QY 359 CTGCTCTTGGCTGATTTTTCAGAGATGGCA-TGGTCTCTGGGATGCTCTGGGCTCTG 417
DB 359 CTGCTCTTGGCTGATTTTTCAGAGATGGCA-TGGTCTCTGGGATGCTCTGGGCTCTG 428
QY 418 TATT-TGTGAATAAGTCTTGTGCAATGCTTCTGCAATGATCCCTTGACACACTTT 476
DB 429 TATTGGGGGATTAAGTCTTGTGCAATGCTTCTGCAATGATCCCTTGACACACTTT 488
QY 477 CCAGCAGATCACTGCAACAGACAGAGAGGAGCGTGTGAAGTATAGACACACCG 536
DB 489 CCAGCAGATCACTGCAACAGACAGAGAGGAGCGTGTGAAGTATAGACACACCG 548
QY 537 ATGTGTGACAAAGATCCGATTTGAGAGCGGTTCACAAACAGTAA-GTGTCTCTGTAC 595
DB 549 ATGTGTGACAAAGATCCGATTTGAGAGCGGTTCACAAACAGTAAAGGTCTCTGTAC 608
QY 596 C-TGGAAAGTGGCTGGAACAACAAGAAACGGCTCTTGGCTGATGCTCCATAGTG 654
DB 609 CTTGAAAGTGGCTGGAACAACAAGAAACGGCTCTTGGCTGATGCTCCATAGTG 668
QY 655 ATTGGAATGAGTGGTGTGATGAGAGCTTCTGAGAGAGAGATGTAGACATC 714
DB 669 ATTGGAATGAGTGGTGTGATGAGAGCTTCTGAGAGAGAGATGTAGACATC 727
QY 715 CTTGCAATTTGTGATGATGTGGCAACAGGCAACAAATTAAGACCAAGATTTAC 774
DB 728 CTTGCAATTTGTGATGATGTGGCAACAGGCAACAAATTAAGACCAAGATTTAC 785
QY 775 CCA 777
DB 786 CAA 788

RESULT 5
LOCUS B1488689
DEFINITION 60302123F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191688 5',
mRNA sequence.
ACCESSION B1488689
VERSION B1488689.1 GI:15327917
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1022)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rc@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM11479 row: f column: 09
High quality sequence stop: 680.
Location/Qualifiers
1..1022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5191688"
/lab_host="DH10B"
/clone_id="NIH_MGC_114"
/note="Organ: brain; Vector: PCMV-SPORE; Site: 1: Noci;
Site 2: EcorV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 75.1%; Score 610.2; DB 12; Length 1022;

Best Local Similarity 97.6%; Pred. No. 4.4e-164;

Matches 684; Conservative 0; Mismatches 8; Indels 9; Gaps 6;

1 CACGTGAGTGGGGATGCGCATGGGCACTAATACCTTCATCAGGAAGATGACACA 60
 12 CACGTGAGTGGGGATGCGCATGGGCACTAATACCTTCATCAGGAAGATGACACA 71
 61 TTATCTCCCATCACTCAAAAGTCTGTCAGAGAGTGAACGCGCAGAGATGATTTAA 120
 72 TTATCTCCCATCACTCAAAAGTCTGTCAGAGAGTGAACGCGCAGAGATGATTTAA 131
 121 AGGTGAAGATGACAGAGTTTCCACCCCTCAAAACCTTGGCTCTTTCTGCAATACAGTC 180
 132 AGGTGAAGATGACAGAGTTTCCACCCCTCAAAACCTTGGCTCTTTCTGCAATACAGTC 191
 181 TGAATGAACCGGATGCTTTTCTTTTCTGCTGGAATAGATCGGAAGAGATACATTT 240
 192 TGAATGAACCGGATGCTTTTCTTTTCTGCTGGAATAGATCGGAAGAGATACATTT 251
 241 TTTTCTTTTATCTGATTAAGAGATGTTGGGAGAGCTCTTGAAGAAAAATTTTAAATG 300
 252 TTTTCTTTTATCTGATTAAGAGATGTTGGGAGAGCTCTTGAAGAAAAATTTTAAATG 311
 301 TGGCAGCAGATGATTTTAAAGATGTTAGATCTTCCATGAGACATATAGATGATCT 360
 312 TGGCAGCAGATGATTTTAAAGATGTTAGATCTTCCATGAGACATATAGATGATCT 371
 361 GCTCTTGGCTGGATTTTTCAGAGATGCAATGCTCTGCGATGCTCGGCTCTGAT 420
 372 GCTCTTGGCTGGATTTTTCAGAGATGCAATGCTCTGCGATGCTCGGCTCTGAT 431
 421 TTGTGATTAATGCTTTGTGCAA-TGCTACTTGGCATGATCCCTTAGACACATTTCCA 479
 432 TTGTGATTAATGCTTTGTGCAA-TGCTACTTGGCATGATCCCTTAGACACATTTCCA 491
 480 GCACATCACTGTCAGACAGACAGAGAGGAGCGTGAAGTATAGACAGACACCGATG 539
 492 GCACATCACTGTCAGACAGACAGAGAGGAGCGTGAAGTATAGACAGACACCGATG 548
 540 TTGTNACAAAGATGCAATTGAGAGAGCGGTCA-CAAAAGTAAAGTGTCTGTACTCTG 598
 549 TTGTNACAAAGATGCAATTGAGAGAGCGGTCA-CAAAAGTAAAGTGTCTGTACTCTG 608
 599 GAAAGATGCTGGA--CAACAAAGAACCGGCTTTCTTGGCTGATGCTCTCATATGAT 656
 609 GAAAGATGCTGGA--CAACAAAGAACCGGCTTTCTTGGCTGATGCTCTCATATGAT 668
 657 TGGGA--ATGCTGCTG--AGATGAGAGCTTGGCTTAAGAG 695
 669 TGGGAACATGCTGCTGCTGAGATGAGAGCTTGGCTTAACAGAG 709

RESULT 6

AKO50001 3258 bp mRNA linear HTC 20-SEP-2003

LOCUS AKO50001 Mus musculus adult male hippocampus cDNA, RIKEN full-length

DEFINITION enriched library, clone: C630045F22 product: unclassified, full

insert sequence.

ACCESSION AKO50001

VERSION AKO50001.1 GI:26093743

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

1 Carrincci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

2 Carrincci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagata, S., Sasaki, N., Carrincci, P.,

Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, K.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the

PANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 The PANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3258)

Adachi, J., Aizawa, K., Akimura, T., Aizawa, T., Bono, H., Carrincci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayatsu, N., Hirose, K., Hirose, T., Hirose, T.,

Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,

Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnishi, N.,

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Takeda, Y., Tanaka, T., Tomaru, A., Toyo, T., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-reeegsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

Fax: 81-45-503-9216)

cdna library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1..3258

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

FEATURES

source

/db_xref="FANTOM_DB:C630045F22"
 /db_xref="MGI:2417712"
 /db_xref="taxon:10090"
 /clone="C630045F22"
 /sex="male"
 /tissue_type="hippocampus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 misc_feature
 1..3258
 /note="unclassifiable"

Query Match 67.9%; Score 551.6; DB 11; Length 3258;
 Best Local Similarity 82.5%; Pred. No. 4e-147;
 Matches 695; Conservative 0; Mismatches 109; Indels 38; Gaps 4;

Db 2 ACTGAGTGGGAGTGTCTCATCGGCACTATAAGTATCTCATGAGAACTGCACAT 61
 31 ACTGAGTGGGAGTGTCTCATCGGCACTATAAGTATCTCATGAGAACTGCACAT 88
 QY 62 TATCTCCCATCACTTCAAAAGTCTCTGAGGAGAGGTGACGCCGAGAGATGATTAA 121
 Db 89 TAGCTCAACATCACTTCAAAAGTCTCTGAGGAGAGGTGACGCCGAGAGATGATTAA 148
 QY 122 GGTGAAATGACAGAGTTTCCACCCCTCAAACTGGCTCTTTTTCGACATACAGTCT 181
 Db 149 GGTGAAATGACAGAGTTTCCACCCCTCAAACTGGCTCTTTTTCGACATACAGTCT 206
 QY 182 GAATGAACCCGATCTCTTTTCTTCTGTAAGTATGATCGAAGAGATGATC----- 236
 Db 207 GAAAGACCCAGCTC---TTCAACATCTGTAAGAGGAGTGGAAAGAAAGCTTGGC 263
 QY 237 -----ATTTTCTTTTAACTCTGTAAGAGATGTTG 270
 Db 264 TTTTCTTTCTTTCTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 323
 QY 271 GGAAGCTCTTGAAGAAAAATTTTAATGTGACAGAGATGATTTTAAAGCTTGA 330
 Db 324 AAGCATTTCTTTGAAATATCTGAACTGTGACAGAGATGATTTTAAAGCTTGA 383
 QY 331 TCTTCCATGACACTAATAGAGTCTGCTCTTGGCTGATTTTTCAGAGATGCA 390
 Db 384 TCTTCCATGACACTAATAGAGTCTGCTCTTGGCTGATTTTTCAGAGATGCA 443
 QY 391 ATGCTCTGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
 Db 444 ATGCTCTGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
 QY 451 TGCCATGATCCCTTACAGACATTTTCCAGAGATGCTGCTGCTGCTGCTGCTGCT 510
 Db 504 TGCCATGATCCCTTACAGACATTTTCCAGAGATGCTGCTGCTGCTGCTGCTGCT 563
 QY 511 ACGTGTGAAGTATGAGGACACACGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
 Db 564 ACGTGTGAAGTATGAGGACACACGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
 QY 571 CAACAGTAAAGTGTCTCTCTCACTGAAAGAGTGTGAGACACAGAACTGGCT 630
 Db 624 CAACAGTAAAGTGTCTCTCTCACTGAAAGAGTGTGAGACACAGAACTGGCT 683
 QY 631 TCTTCTGATGCT 690
 Db 684 TCTTCTGATGCT 743
 QY 691 GAAAGAGAGATGTAAGACACTCCCTGACATTTCTGATGATGATGATGATGATG 756
 Db 744 GAAAGAGAGATGTAAGACACTCCCTGACATTTCTGATGATGATGATGATGATG 803
 QY 751 AAAATTAGACACAGAGATTCACCCAGAACTTAAAGAGATTTTGTAGTAAAG 810
 Db 804 AAGATTAAAGATCAAGATTTCACCCAGAACTTAAAGAGATTTTGTATATAAATAG 863
 QY 811 AA 812

Db 864 GA 865

RESULT 7
 CCR001752
 LOCUS
 DEFINITION AGNCOURT 16385450 NIH MGC 227 Homo sapiens cDNA clone
 IMAGE:30718008 5', mRNA sequence.
 ACCESSION CCR001752
 VERSION CCR001752.1 GI:38527786
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 672) 672 bp mRNA linear EST 26-NOV-2003
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: dsgerha@mail.nih.gov
 Tissue Procurement: Professor Miklos Palokovits
 cDNA Library Preparation: Michael Brownstein / Ted Usdin
 Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: NDCM260 row: j column: 01
 High quality sequence stop: 579.
 Location/Qualifiers

FEATURES

1..672
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30718008"
 /tissue_type="Bull tissue from Human spinal cord"
 /lab_host="DH10B Tona"
 /clone_lib="NIH MGC 227"
 /note="Organ: brain/CNS; Vector: pDNR-LIB; Site 1: Sfil
 (directional); Site 2: Sfil (directional); Library is
 oligo-dT primed and directionally cloned; 5' and 3'
 adaptors were used in cloning as follows:
 5'-AGCAGGTGATCAACGACAGAGTGGCTTACGCGCGG-3'
 5'-ATTCTAGAGCCGAGGCGCGGACATG-d(T)30N-IN-3' Full-length
 enriched library was constructed using the clonech
 creator SMART kit and size-selected for >0.5kb with an
 average insert size of 1.3kb Library created in the
 laboratory of Jonathan Kuo and Ted Usdin."

ORIGIN

Query Match 62.9%; Score 511; DB 14; Length 672;
 Best Local Similarity 100.0%; Pred. No. 1.4e-15;
 Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 91 AGCAGAGTGAACGCCAGAGATGATTTAAAGTGAATAAGCAAGTTTCCACCCCTCA 150
 Db 4 AGCAGAGTGAACGCCAGAGATGATTTAAAGTGAATAAGCAAGTTTCCACCCCTCA 63
 QY 151 AACCTTGAGCTCTTTTCTGACATACAGTCTGAATGAACCCGATGCTTTTTTACTG 210
 Db 64 AACCTTGAGCTCTTTTCTGACATACAGTCTGAATGAACCCGATGCTTTTTTACTG 123
 QY 211 TGAATTAAGATCGAAGAGATTAATTTTTTTTAACTGATTAAGAAAGATTTGTTG 270
 Db 124 TGAATTAAGATCGAAGAGATTAATTTTTTTTAACTGATTAAGAAAGATTTGTTG 183
 QY 271 GGAAGCTCTTTGAAAAAATTTTAATTTGTGACAGATGATTTTAAAAAGTTTGA 330


```

FEATURES             SOURCE
SOURCE              1..768

REFERENCE
AUTHORS             Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Butt,D.W., Bosch,E.,
TITLE               Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
JOURNAL            A Comprehensive Collection of Chicken CDNAS
MEDLINE            Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED             22335534
COMMENT             12445392
                  Contact: Simon Hubbard
                  Department of Biomolecular Sciences
                  University of Manchester Institute of Science and Technology
                  (UMIST)
                  PO Box 88, Manchester, M60 1QD, UK
                  Tel: 01612008930
                  Fax: 01612360409
                  Email: Simon.Hubbard@umist.ac.uk.
                  Location/Qualifiers
                  1..768

```

ORIGIN	[5'/aattccttttttcgataccgg999ctgcacgc/']
Query Match	58.5%; Score 475.4; DB 13; Length 768;
Best Local Similarity	82.3%; Prid. No. 2,4e-125;
Matches	572; Conservative 0; Mismatches 116; Indels 7; Gaps 2;
Qy	98 GGTGACCGCAGAGATGATATTAAAGGGAATAATGACAAGTTTCCACCCCTCAAACTGTG 157
Db	10 GGAGGCGGGCCCCGACGAGGTGAAGTGAATGAAAGATTTCACCCCTCAAACTCTCG 69
Qy	158 GCTCTTTTTCGACATATACAGTCTGATG--AACCGATGCTTTTTTTTTTACTGTGAA 215
Db	70 GCGACATTTTCGAGGCGAGTGCCTGAGAGCTGAGATGTGCTTTTACTGTGGAAACAGA 129
Qy	216 ATGAGATCGGAAGAGTAACTTTTTTTTTTATCTCTATAAAGAAATTGTGGGAAG 275
Db	130 TTGTGAACAGAGATTTTTTTTTTGTCTGCTGATCTATTAAGAAAGATATTGGGAAG 189
Qy	276 CTCTTGGAAAAAATTTTAAATTTGAGCAGATGAGATTTTAAAGTTAGATCTT 335
Db	190 CTATT-----AAAAATCCTTATTTGTGTGCAATGATTTTAAAAAGTTAGCTTT 244
Qy	336 CCAATGACACTAATATAGTACTCTGCTCTGTGCTGATTTTTCAGAGATGGCAATGAT 395
Db	245 CCAATGAACTAATATAGTCTCTGCTCTTGGCTGATTAATCTCAGAGAAATGGACGTGT 304
Qy	396 CTCTGCGATGTCCTGGGTCTGTATTTGTGATGATGCTTGTGCAATGCTACTCTGCA 455
Db	305 CTCTGCAATGTCTGGGTCCGTGATTTGTGAGATGATGCTGTGCAATGCTGCTGTCA 364
Qy	456 TGGATTCCTTTAGACACTTTTCCAGCAGCATCTCTGCAAGACAGAAAGAGGAGAGCTG 515
Db	365 GGGATCCCTTCATCAGACTTTTCCAGCAGCATCTCTGCAAGACAGAAAGAGGAGAGCTG 424
Qy	516 TGAATGATPACAGACACACGATGTTGTAAACAAGATTCGATTTGAGAGCGGTCAACAAC 575
Db	425 TGAATGATPACAGACACAGATGCTGTATTAAGATTCGATTTGAGAGAGATCAACAAC 484
Qy	576 AGTAAAGTTCCTGTCTACCTGGAAGATGCTGGAACACAGAAACCGGCTTTCTTG 635
Db	485 AGTAAAGTTCCTGTCTACCTGGAAGATGCTGGAACACAGAAACAGACTTTCTTG 544
Qy	636 CCGTCAATGCTTCATGATGATTTGGGAATGTGTGTGAGATGAGACCTTGCTTGAAAG 695
Db	545 TGTATATCTTCATATGATGATTTGGGAATGTGTGTGAGATGAGACCTTGCTTGAAAG 604
Qy	696 AATAAATGTAGAACAACCTCCCTGCAATTTGTGATGAGATGTGGCAAGCAACAAAT 755
Db	605 AAGAAGATGTAGACGTTGCTGTATTTGTGATGAGATGTGTGCACTTGGCAATTAAT 664
Qy	756 TAAGACCAAGAAATTCACCCAGAAACCTTAACAGA 790
Db	665 CAAGACCAAGAAATTCACCCAGAAACCTTAACAGA 699
RESULT 9	
BF652347	506 bp MRNA linear EST 25-APR-2001
LOCUS	BF652347
DEFINITION	275877 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION	BF652347
VERSION	BF652347.1 GI:11917477
KEYWORDS	EST.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
	Bovidae; Bovinae; Bos.
	1 (bases 1 to 506)
	Smith,T.P.L., Groves,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
	Cassas,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C.,
	Bennett,G.L., Heaton,M.P., Liegreid,W.W., Rohrer,G.A.,
	Chitko-McKown,C.G., Petrea,G., Holt,I., Karamycheva,S., Liang,F.,

TITLE Quackenbush, J. and Keefe, J.M.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 PUBMED 11282978
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 16
 and -minmatch 12 options.
 PCR primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCOCAGTCAGACG
 Plate: 65 row: M column: 8
 Seq primer: ATTAGTGACACTATAG.
 Location/Qualifiers
 1..506

FEATURES

source

/organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_id="MARC 3BOV"
 /note="Vector: pCMV SPORT6; site 1: NotI; site 2: SalI;
 library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."

ORIGIN

Query Match 54.5%; Score 442.2; DB 10; Length 506;
 Best Local Similarity 93.3%; Pred. No. 7.7e-116;
 Matches 473; Conservative 0; Mismatches 33; Indels 1; Gaps 1;
 Oy 249 ATCTGATTAAGAAAGATGTTGGGAAGCTCTTGAAGAAAATTTAATTGTGGACAG 308
 Db 1 ATCTGATTAAGAAAGATGTTGGGAAGCTCTTGAAGAAAATTTAATTGTGGACAG 59
 Oy 309 ATGATTTTAAAGTGTAGATCTTTTCCAAAGAACTAAATAGATCTGCTCTGG 368
 Db 60 ATGATTTTAAAGTGTAGATCTTTTCCAAAGAACTAAATAGATCTGCTCTGG 119
 Oy 369 CTGATTTTTCAGAGATGGCAATGCTCTGCGCATGCTCGGGTCTGTATTTGTGGAT 428
 Db 120 CTGATTTTTCAGAGATGGCAATGCTCTGCGCATGCTCGGGTCTGTATTTGTGGAT 179
 Oy 429 AAGTCTGTGGAAGTACTCTGCGCATGATCCCTTCAGACACTTTCAGACAGATCA 488
 Db 180 AAGTCTGTGGAAGTACTCTGCGCATGATCCCTTCAGACACTTTCAGACAGATCA 239
 Oy 489 CTTGACAGACAGAGAGAGAGAGCTGTGAAGTATAGACACACACCATGTTTAACAA 548
 Db 240 CTTGACAGACAGAGAGAGAGAGCTGTGAAGTATAGACACACCATGTTTAACAA 299
 Oy 549 GAATGTCATTGAGAGCGGTCAAAACAGTAAAGTCTTCTGTTCACTTGGAAAAAGTGGC 608
 Db 300 AAACCGCATTTGAGAGCGGTGCAAAACAGTCAAGTCTTCTGTTCACTTGGAAAAAGTGGC 359
 Oy 609 TGGAAACAACAAAGAAACCGGCTTCTGGCTGATCCCTCCATAGATTTGGGAATGGTG 668
 Db 360 TGGAAACAACAAAGAAACCGGCTTCTGGCTGATCCCTCCATAGATTTGGGAATGGTG 419
 Oy 669 GTGTAGATGAGAGCTTGGCTTGAAGAGAAAGATGTAAGACATCCCTGATATTTGG 728
 Db 420 GTGTAGATGAGAGCTTGTCTTGAAGAGAAAGATGTAAGACATCCCTGATATTTGG 479
 Oy 729 ATGATGTGCGCAACAGGCAACAAAT 755
 Db 480 ATGATGTGCTACAGGCAATTAAT 506

RESULT 10
 LOCUS BU615139 706 bp mRNA linear EST 20-FEB-2003
 DEFINITION UI-M-FRO-cbe-1-08-0-UI-21 NIH BMAP FRO Mus musculus cDNA clone
 UI-M-FRO-cbe-1-08-0-UI 5', mRNA sequence.
 ACCESSION BU615139
 VERSION BU615139.1 GI:22281354
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 706)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@iowa.edu
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES

source

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-FRO-cbe-1-08-0-UI"
 /tissue_type="whole brain"
 /dev_stage="Embryo 13.5, 14.5, 16.5, 17.5 dpc"
 /lab_host="DH10B (TI phase resistant)"
 /clone_id="NIH-BMAP_FRO"
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
 program coordinator."

ORIGIN

Query Match 54.3%; Score 440.8; DB 13; Length 706;
 Best Local Similarity 81.2%; Pred. No. 2.1e-115;
 Matches 575; Conservative 0; Mismatches 94; Indels 39; Gaps 4;
 Oy 23 CGGCACTTAATTAACGATCTCATGACGAACTGACACTTATCTCCCATCAAGTCAAG 82
 Db 6 CGTNGGCTATTAACGCTCTTCTATCGGAACCT--ACATTAGTCAACATCACTTCAAG 63
 Oy 83 GTCTGTCAGGAGAGAGTGAAGGAGAGATTTAAAGTGAAGTGAAGTGAAGTTC 142
 Db 64 GTCTGTCAGAGAGAGTGAAGGAGAGATTTAAAGTGAAGTGAAGTGAAGTTC 123
 Oy 143 ACCCTCAAACTTGGCTCTTTTCTGAACATPACGCTGAATGAACCCGATGCTTTT 202

Db 124 ACCCTCAAACTGCTTCCTTCCTGACATACAG--TGAACGAGCCCAACGTC--TT 178
 QY 203 TTTTACGTGGAAATAGGATGCGAAGAGATGAC----- 226
 Db 179 CACAACTGTGAAACGGGACCTGGAAGAGAAAGCTTGCTTTTCTTTCTTTCTTT 238
 QY 237 -----ATTTTCTTTTAACTCTGATAAAGAAATGTTGGAAAGCTTTGAAAAAAA 290
 Db 239 TCTTTTCTTTTCTTTTCAATCTTTAAATAAGAAATGCTGGGAAGATCTCTTAAAAAT 298
 QY 291 TTTTAAATTTGGACAGATGATTTTAAAAAGTGTGATCTTTCCATGAACTAAT 350
 Db 299 CTCAGAACTGTGGACAGATGATTTTAAAAAGTGTGATCTTTCCATGAACTAAG 358
 QY 351 AGAGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 410
 Db 359 AGGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
 QY 411 GATCCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 470
 Db 419 GGCCTGATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 478
 QY 471 CACTTTCAGACGATCACTGACAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 530
 Db 479 CACCTTCAGAGAGATCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
 QY 531 ACACGATGTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 590
 Db 539 CCACAGGTGTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 598
 QY 591 TCTACCTGAAAGAGGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCT 650
 Db 599 TTTACCTGGAAGAGGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCT 658
 QY 651 AGTGAATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 698
 Db 659 AGTGAATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 706

RESULT 11
 AL533965 975 bp mRNA linear EST 12-MAY-2003
 LOCUS AL533965 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
 DEFINITION CS0DF003YN3 5-PRIME, mRNA sequence.
 ACCESSION AL533965
 VERSION AL533965
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 JOURNAL 1 (bases 1 to 975)
 COMMENT Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 13, 2001 this sequence version replaced gi:12797458.
 CONTACT Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 2511.f
 Contact: Feng Liang Email: fliang@lifetech.com URL:
 http://fulllength.invitrogen.com/InvitrogenCorporation1600
 Faraday Avenue Genoscope Sequence ID: CS0DF003G12QPL.

FEATURES

source
 1..975
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DF003YN3"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"

ORIGIN
 Query Match 53.5%; Score 434.8; DB 9; Length 975;
 Best Local Similarity 95.6%; Pred. No. 1.2e-113;
 Matches 436; Conservative 9; Mismatches 11; Indels 0; Gaps 0;
 /clone_lib="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
 was primed with a NotI-oligo (dT) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."

QY 357 CTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 416
 Db 411 CTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
 QY 417 GTATTTGTGATTAAGCTGCTTGTGCAATGCTACTTGTCCATGATCCCTTACACACTTT 476
 Db 471 GTATTTGTGATTAAGCTGCTTGTGCAATGCTACTTGTCCATGATCCCTTACACACTTT 530
 QY 477 CCAGCAGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536
 Db 531 CCAGCAGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
 QY 537 ATGTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 596
 Db 591 ATGTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
 QY 597 TGGAAAGTGGCTGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 656
 Db 651 TGGAAAGTGGCTGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 710
 QY 657 TGGAAAGTGGCTGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 716
 Db 711 TGGAAAGTGGCTGCTGGAACAAGAAACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 770
 QY 717 TGAACAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 776
 Db 771 TGAACAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 830
 QY 777 AAGAACTTAACAGAGCAATTTGTGATTAAGAA 812
 Db 831 AAGAACTTAACAGAGCAATTTGTGATTAAGAA 866

RESULT 12
 H23443 468 bp mRNA linear EST 06-JUL-1995
 LOCUS H23443
 DEFINITION ym52909.r1 Soares infant brain INTB Homo sapiens cDNA clone
 ACCESSION H23443
 VERSION H23443
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 JOURNAL 1 (bases 1 to 468)
 COMMENT Full-length cDNA libraries and normalization
 Unpublished (1995)
 CONTACT Wilson RK
 The Wash-Merck EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 948
 High quality sequence stops: 360

Best Local Similarity 79.3%; Pred. No. 1.4e-100;
Matches 526; Conservative 0; Mismatches 99; Indels 38; Gaps 4;

```

QY 2 ACTGAGTGGGATGTCATCGGCACTAATCTCATCAGGAAGTGCACAT 61
Db 31 ACTGAGTGGGATGTCATCGGCACTAATCTCATCAGGAAGTGCACAT 88
QY 62 TATCTCCCATCTCTCAAGGTCCTGTCAGGCGAGGTGACGCCAGAGATTGTTAA 121
Db 89 TAGCTCAACCATCACTTCAAGGTCCTGTCAGGCGAGGTGACGCCAGAGATTGTTAA 148
QY 122 GGTGAAATGACAGGTTTCAACCCCTTAACCTGGCTCTCTTTGACAAATCACTCT 181
Db 149 GGTGAAATGACAGGTTTCAACCCCTTAACCTGGCTCTCTTTGACAAATCACTCT 206
QY 182 GAATGAACCCGATGCTTTTCTAGTGAATAAGATCGGAGAGATAAC----- 236
Db 207 GAAGAGCCCAAGTC--TTCAACTGTGGAACGGGATCGAAGAGAAAGCTTGGCC 263
QY 237 -----ATTTTTCTTTTAACTGTAATAAGAGATTGTTG 270
Db 264 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 323
QY 271 GGAAGCTCTTTGAAAAAATTTTAATTGTCGACAGATGATTTTAAAGTGTGA 330
Db 324 AAGCATCTCTTTGAAAAATCTCAGACACTGTGGCAAGATGATTTTAAAGTGTGAC 383
QY 331 TCTTCCATGAACATAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 390
Db 394 TCTTCCATGAACATAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443
QY 391 ATGCTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
Db 444 ATGCTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
QY 451 TGCCATGATCCCTTCAGACACTTTTCAGACAGATCACTGTCACAGACCAAGAGAGG 510
Db 504 TGCCATGATCCCTTCAGACACTTTTCAGACAGATCACTGTCACAGACCAAGAGAGG 563
QY 511 AGCTGTGAAGTATGACAGACACCGAGTGTGTAACTGATGATGAGAGAGGTC 570
Db 564 AGCTGTGAAGTATGACAGACACCGAGTGTGTAACTGATGATGAGAGAGGTC 623
QY 571 CAACAGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630
Db 624 CAACAGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 683
QY 631 TCT 633
Db 684 CCT 686

```

RESULT 14
AK049880
LOCUS
DEFINITION
AK049880 2412 bp mRNA linear HTC 20-SEP-2003
Mus musculus adult male hippocampus cDNA, RIKEN full-length
enriched library, clone:0630007B19 product:unclassified, full
insert sequence.
ACCESSION
AK049880
VERSION
AK049880.1 GI:26093679
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
MEDLINE
PUBMED
10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, K., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2412)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Inokami, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Nunuma, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, "The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location:Qualifiers

FEATURES

source

1. .2412
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:C630007B19"
/db_xref="WGI:2417405"
/db_xref="taxon:10090"
/clone="C630007B19"
/sex="male"
/tissue_type="hippocampus"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature
1. .2412

Db 734 AATGCTGTGGAATGAGCCTGCTAGAGGAGAGGATGTACGACCGTTGCTGA 793
Qy 720 CAATCTGATG 732
Db 794 TAACTGATG 806

Search completed: April 9, 2004, 15:45:13
Job time : 2566 secs